IRE Transactions
on ELECTRONIC COMPUTERS

Volume EC-6 **SEPTEMBER, 1957** **Number 3**

Published Quarterly

TABLE OF CONTENTS

CONTRIBUTIONS

CORRESPONDENCE

PUBLISHED BY THE
Professional Group on ELECTRONIC COMPUTERS

IRE Transactions® on Electronic Computers

Published by the Institute of Radio Engineers, Inc., for the Professional Group on Electronic Computers at 1 East 79th Street, New York 21, N.Y. Responsibility for the contents rests upon the authors and not upon the IRE, the Group, or its members. Price per copy: IRE PGEC members, $1.25; IRE members, $1.85, nonmembers, $3.75. Yearly subscriptions rate: nonmembers, $17.00; colleges and public libraries, $12.75. Address requests to The Institute of Radio Engineers, 1 East 79th Street, N.Y. 21, N.Y.

Notice to Authors: Address all papers and editorial correspondence to J. P. Nash, Missile Systems Division, Lockheed Aircraft Corp., Sunnyvale, Calif. To avoid delay, 3 copies of papers and figures should be submitted, together with the originals of the figures which will be returned on request. All material will be returned if a paper is not accepted.

Analog Computer Applications in Predictor Design*

M. R. BATES†, D. H. BOCK†, AND F. D. POWELL†

Summary—The theory of least mean-squared-error filtering and prediction of statistical time series, developed by Kolmogorov,[1] Wiener,[2] Bode and Shannon,[3] and others,[4] has recently found applications in a variety of control systems.

The design of optimum (least mean squared-error) predictors is based on measurement or calculation of power spectra. The present paper describes equipment for measuring power spectra at very low frequencies. It is shown that analog circuits having the transfer functions of exact stagger-tuned triples, with flat response in the pass band and good skirt selectivity, are easily designed and utilized in conjunction with magnetic tape recordings.

Although the optimum predictor transfer function can be computed directly from the power spectrum of the signal to be predicted, the computation is complicated and a checking procedure is desirable. An analog computation which permits such a check is described.

Since the power spectrum defines the optimum predictor, it determines the mean-squared prediction error as well. However, if a nonoptimum network is arbitrarily selected for use as a predictor, it becomes desirable to determine the incremental error that results from use of this network. A simple analog computation which determines the error of the optimum predictor and the incremental error of a nonoptimum predictor is described. This analog technique is a specific adaptation to prediction of the impulsive response techniques of Bennett[5] and Laning and Battin.[6,7]

The ultimate test of a predictor, the comparison of its observed error and its theoretical error, gives an indirect check on the predictor design and can again be performed on analog computers.

Introduction

CONSIDERABLE interest has been displayed in the optimum prediction problem ever since the publication of the pioneering work of Kolmogorov[1] in Russia and Wiener[2] in the United States. While the theoretical aspects of this work have been described quite thoroughly, little has been written about practical procedures in the design and testing of prediction networks. The prediction problem of interest here concerns the selection of a device to estimate the instantaneous magnitude of a signal at a time α seconds in the future from a linear operation on the past of the signal. The device is chosen to minimize the mean-squared error between the actual value of the signal at a time $t+\alpha$ and the value predicted for that instant. (We are not concerned here with a forecasting process which involves an estimate of the future statistical behavior of a signal.) The input signal is assumed to be random rather than an analytically specified signal such as a periodic signal, a polynomial, or a damped sine wave. The fact that the signal is not analytically defined does not preclude an estimation of its future if certain statistical parameters are specified. It can be shown that the optimum time-invariant linear predictor (yielding the smallest mean-squared error) is completely determined by the frequency content of the signal expressed as a power spectrum, or indirectly as a correlation function. Filter techniques for measuring power spectra directly will be presented.

As might be expected, a signal is most predictable if its power spectrum is most sharply concentrated. This point is illustrated in the Appendix, where the mean-squared error of prediction is plotted against prediction interval for a family of band-pass power spectra with varying Q's.

This calculation of the mean-squared error, that will result when the optimum predictor is used to estimate the future value of a signal with a given spectrum, was performed on an analog computer. The technique is similar to the impulsive response technique described by Bennett[5] and Laning and Battin[6,7] for calculating the mean-squared output of a network excited by filtered random noise.

In a class of problems that have been of especial interest, the power spectra may be characterized by a group of parameters such as central frequency, half-power bandwidth, dc level, and high-frequency cutoff rate. In one case, short-term power spectra were measured over periods of one-half hour to six hours and it was observed that these spectra slowly changed with time. It was thought desirable to design a predictor corresponding to some average values of the spectral parameters, and then determine the incremental error that results from using the designed predictor for perturbed power spectra. For the power spectra considered, it was determined that the central frequency was the most critical parameter in designing a predictor to approximate the behavior of the signal about half a period in advance. This eventually led to the design of a time-varying predictor which measures the zero crossings of a signal and adjusts a transfer function to respond to the measured central frequency.

* Manuscript received by the PGEC, May 1, 1956. This research was supported by the U. S. Navy.
† Bell Aircraft Corp., Buffalo, N. Y.

[1] A. N. Kolmogorov, "Interpolirovanie i ekstrapolirovanie stacionarnykh slučajnykh posledovatel'nostej," *Izvest. Akad. Nauk SSSR, Ser. Matematicheskaia*, vol. 5, pp. 3–14; 1941.

[2] N. Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series," John Wiley and Sons, Inc., New York, N. Y.; 1949.

[3] H. W. Bode and C. E. Shannon, "A simplified derivation of linear least square smoothing and prediction theory," Proc. IRE, vol. 38, pp. 417–425; February, 1950.

[4] N. Levinson, "A heuristic exposition of Wiener's mathematical theory of prediction and filtering," *J. Math. and Phys.*, vol. XXVI, pp. 110–119; July, 1947.

[5] R. R. Bennett, "Analog computing applied to noise studies," Proc. IRE, vol. 41, pp. 1509–1513; October, 1953.

[6] J. H. Laning, Jr., and R. H. Battin, "An application of analog computers to problems of statistical analysis," Cyclone Symp. II, Reeves Instrument Corp., p. 79; 1952. Also IRE Trans., Vol. CT-2, p. 44; March, 1955.

[7] J. H. Laning, Jr., and R. H. Battin, "Random Processes in Automatic Control," McGraw-Hill Book Co., Inc., New York, N. Y.; 1956.

The analog technique that was used to calculate the incremental error resulting from use of a nonoptimum predictor will be explained. A theoretical justification of this technique will be presented, and some specific calculations will be shown. It should be noted that the present paper is confined to the problem of pure prediction (in the absence of noise). Similar techniques can be applied to the problem of filtering or combined prediction and filtering, but space and time limitations prevent significant discussion of these techniques in the present paper.

Power Spectral Measurement

General

Application of Wiener's technique to the prediction of a particular random signal requires an accurate determination of the power spectrum of that signal. Most of the previously published work (*e.g.*, Cohen[8]) obtains the power spectrum indirectly as the Fourier transform of the autocorrelation function. An adequate measurement of the power spectrum of a given signal may also be obtained through the use of appropriate band-pass filters. An explanation of this direct technique of power spectral measurement is presented. The technique is illustrated by its application to the measurement of the spectrum of the previously mentioned half-hour magnetic tape record.

Data Sources

The signals of interest here, such as airplane oscillations, gusts, sea motion, and ship motion, are very low-frequency phenomena. They are most conveniently handled on an analog computer if they are available in the form of slowly varying dc signals. This may be achieved by fm recording of the measured data on magnetic tapes. Simultaneous recording of several signals is achieved by modulating a series of subcarriers, mixing and recording on magnetic tapes. Available subcarrier frequencies include 1.3, 1.7, 2.3, 3.0, 3.9 kc, etc., up to 40 kc. When the signal is required, the tape is played back and the information is extracted by a combination of band-pass filters and discriminators. Ampex now markets an fm recording package that is capable of recording up to six hours of 14-channel data on a $10\frac{1}{2}$-inch reel of tape.

For illustrative purposes, a half-hour magnetic tape record of a random low-frequency phenomenon was secured.

These data were recorded using older Ampex equipment and a 3.9-kc subcarrier oscillator. A portion of the demodulated signal was recorded on a Sanborn recorder and is shown in Fig. 1. All techniques discussed in the present paper will be illustrated by their application to these recorded data.

[8] R. Cohen, "Some Analytical and Practical Aspects of Wiener's Theory of Prediction," M.I.T. Res. Lab. Electronics Rep. No. 69; 1948.

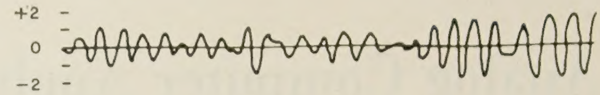


Fig. 1—Sample of a random input signal.

Theory

The power spectrum of a random signal, $f(t)$, may be defined[9] by

$$\Phi_f(\omega) = \lim_{T\to\infty} \frac{1}{2T}\left| \int_{-T}^{T} f(t)e^{-i\omega t}dt \right|^2. \quad (1)$$

It is readily shown that the power spectrum, $\Phi_0(\omega)$, of the output of a linear network when excited by a random input, $f(t)$, may be written as

$$\Phi_0(\omega) = \left| k^2(\omega) \right| \Phi_f(\omega), \quad (2)$$

where $\Phi_f(\omega)$ is the power spectrum of the input signal and $k(\omega)$ is the frequency domain transfer function of the network. See, for example, Bennett.[5] In addition, the mean-squared magnitude of a signal is related to the power spectrum of the signal by

$$\overline{f^2(t)} = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f^2(t)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_f(\omega)d\omega. \quad (3)$$

These properties of the power spectrum justify the use of the mean-squared output of a narrow band-pass filter as a measure of the magnitude of the power spectrum over the pass band of the filter. If a random signal $f(t)$, having a power spectrum $\Phi_f(\omega)$, is supplied as an input to a sufficiently narrow band-pass filter then the filter output has a spectrum whose power is largely confined to the pass band of the filter (Fig. 2). The mean-squared filter output is proportional to the shaded area under the curve (Fig. 2) and is, therefore, a measure of the average spectral content of $f(t)$ over the pass band of the filter. Ideally, the filter should have constant (unit) gain at all frequencies within the pass band, and zero gain (infinite attenuation) elsewhere. Such behavior is not physically realizable, but it is possible to approximate this behavior to any desired accuracy.

An indication of the performance of physically realizable filters is presented in Fig. 3, where the squared gain of three band-pass filters is compared. These filters are identified as a simple band-pass filter, whose squared gain is given by

$$\left| k_1^2(\omega) \right| = \frac{\omega^2/\omega_0^2Q^2}{(1-\omega^2/\omega_0^2)^2 + \omega^2/\omega_0^2Q^2}, \quad (4)$$

a cascaded identical n-tuple, whose squared gain is

$$\left| k_2^2(\omega) \right| = \left[\frac{\omega^2/\omega_0^2Q^2(2^{1/n}-1)}{(1-\omega^2/\omega_0^2)^2 + \omega^2/\omega_0^2Q^2(2^{1/n}-1)} \right]^n \quad (5)$$

and an exact staggered n-tuple, whose squared gain is

[9] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 23, pp. 282–332; July, 1944. Also vol. 24, pp. 46–157; January, 1945.

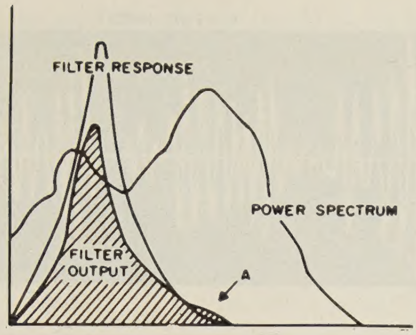


Fig. 2—Filter input-output comparison (frequency domain). Note: The output of a power spectral filter is mostly confined to the pass band of the filter. With poor choice of filter characteristics, however, off-band peaks may cause error (A above).

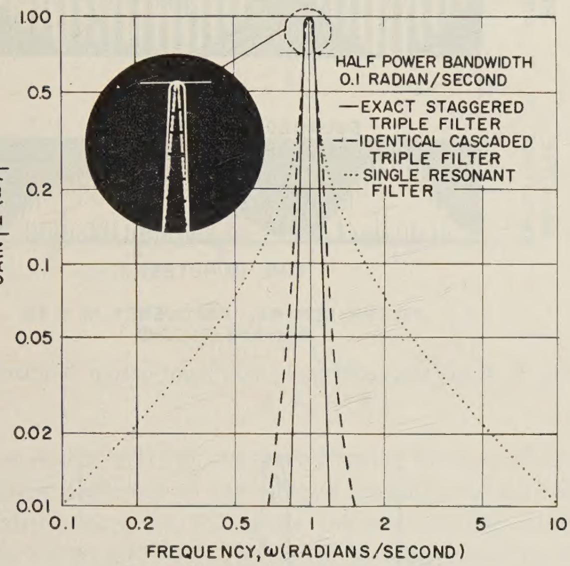


Fig. 3—Gains of various band-pass filters.

$$\left| k_3{}^2(\omega) \right| = \frac{(\omega^2/\omega_0{}^2Q^2)^n}{(1 - \omega^2/\omega_0{}^2)^{2n} + (\omega^2/\omega_0{}^2Q^2)^n}. \tag{6}$$

Magnitudes of $n=3$ and $Q=10$ were considered. In each case, Q represents the ratio of central frequency, ω_0, to half-power bandwidth. All plots were made against a nondimensional frequency, ω/ω_0.

It will be noted that the exact stagger-tuned filter has the flattest characteristic within the pass band as well as the sharpest rejection outside this pass band. These are both desirable characteristics in a filter. While such stagger-tuned filters have been used in radar work, this is believed to be a new application to frequencies below 1 cps.

A quantitative measure of performance may be obtained by comparing the filter response to each of a pair of specified signals. The first signal has its entire frequency content confined to the half-power bandwidth of the filter, and is flat over this frequency band. The second signal has a completely flat spectrum. The ratio of the mean-squared filter output resulting from the first signal to the mean-squared filter output resulting from the second indicates the efficacy of the filter in rejecting the undesired portion of a signal. This ratio is unity for an ideal filter. It has been calculated for a simple band-pass filter, a cascaded identical triplet, and an exact staggered triplet, with results of 0.500, 0.694, and 0.863, respectively. In the case of an exact staggered n-tuple, this ratio is given by

$$R(n) = \frac{1}{2} + \frac{1}{\pi} \sin \frac{\pi}{2n} \sum_{k=1}^{[n/2]} \cos \frac{\pi(2k-1)}{2n} \ln \left[\frac{1 + \cos \dfrac{\pi(2k-1)}{2n}}{1 - \cos \dfrac{\pi(2k-1)}{2n}} \right] \tag{7}$$

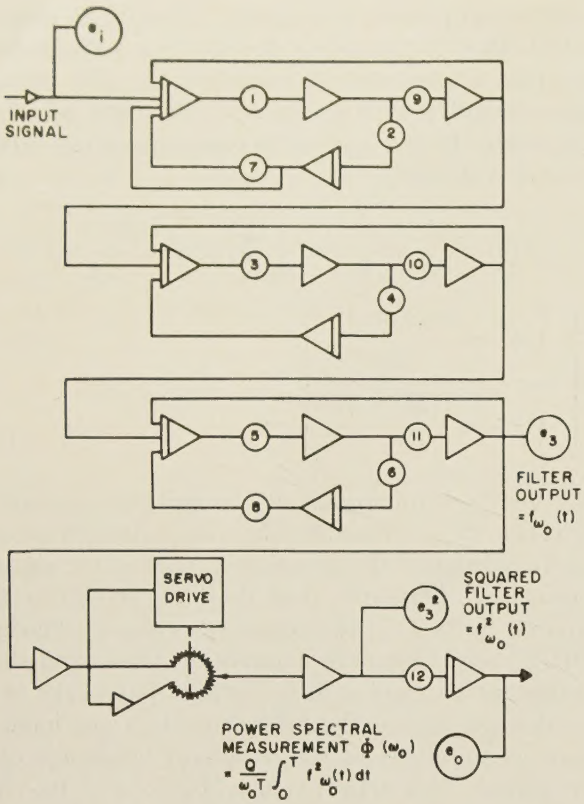
which increases monotonically to unity as n increases.

Selection of an adequate filter must depend on some prior knowledge of the power spectrum of the signal to be measured. Thus, the Q of the filter should be large compared to the Q of the power spectrum of the signal. Similarly, signal rejection outside the pass band should be sufficient to yield a filter output that peaks at the natural frequency of the filter, and does not have significant components at the dominant frequency of the input signal. This rejection may be secured by choice of sufficiently large Q, and selection of exact stagger tuning with an adequately large n, although the size of these quantities is limited by the length of the available data sample. The choice of filter parameters should be validated by appropriate monitoring of the filter outputs. Thus, the central frequency of the filter output should fall within the pass band, and the ring-out time should not be excessive.

Analog Technique

A simple analog circuit for an exact staggered triplet is shown in Fig. 4. It will be noticed that this triple stagger-tuned filter requires nine amplifiers (although twelve are utilized to simplify pot setting). In addition, a squaring device and a low-pass filter (for averaging) are required. Determination of a power spectrum requires repeated runs with the same data or use of a large number of analog computers.

The analog technique for measuring power spectra was employed to determine the power spectrum of the previously mentioned signal. The 3.9-kc tape recorder output was demodulated and supplied as an input to the analog circuit shown in Fig. 4. A Q of 10 was selected, and the mean-squared filter output was evaluated at each of ten different frequencies. A comparison of the filter output and the square of the filter output with the filter input, when the filter is tuned to a central frequency of 0.660 radians per second, is presented in Fig. 5. For the filters discussed here, the response to a unit white noise has a mean-squared value which is equal to ω_0/Q times a constant that depends solely on the form of the filter. For the exact staggered triple, this constant is



POTENTIOMETER	1	2	3	4	5	6
SETTING	ω_0	ω_0	ω_0	ω_0	ω_0	ω_0

POTENTIOMETER	7	8	9	10	11	12
SETTING	0.091	0.917	0.052	0.100	0.048	$\frac{Q}{\omega_0 T}$

Fig. 4—Power spectral analyzer.

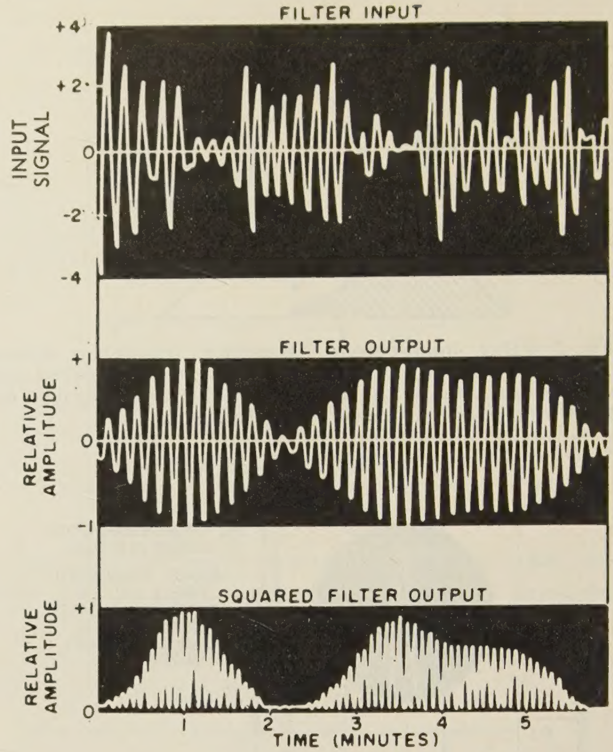
$$\frac{\pi}{6} + \frac{\sqrt{3}}{6} \ln (2 + \sqrt{3}).$$

Thus, the mean-squared value of the filter output must be divided by 0.90377 ω_0/Q before being interpreted as a point of the power spectrum. The measured points of the power spectrum are plotted in Fig. 6.

As indicated earlier, this analog technique of spectral measurement requires repeated runs at different filter frequencies with the same input data supplied to the filters. This obviously requires a large expenditure of time or use of a large quantity of equipment.

Simplified Spectral Analyzer Equipment

In order to speed the measurement, it has proven necessary to design special band-pass filters, employing modified dc amplifiers. The amplifier uses a dual triode on the input with cathode coupling to provide subtraction. The two grids are then available for positive and negative feedback, respectively. Many commercially available dc amplifiers may be modified for the present purpose by disconnecting the chopper stabilizer. The remainder of the circuit is provided for amplification,



FILTER CENTRAL FREQUENCY IS 0.66 RADIANS/SECOND

Fig. 5—Exact staggered triple filter input-output comparison.

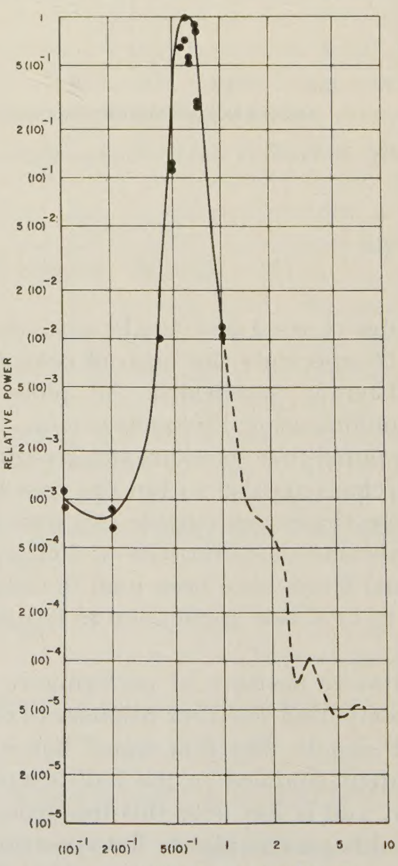


Fig. 6—Measured power spectrum of sample signal.

except that a cathode follower is added for isolation between the constituent band-pass filters.

The desired filter characteristic is secured by employing resistive negative feedback together with a Wien bridge in the positive feedback circuit (see Fig. 7). It may be shown that the transfer function of the circuit is given by

$$\frac{F(e_0)}{F(e_1)} = -\frac{k'j\omega}{-\omega^2 + \omega_0{}^2 + j\omega\omega_0/Q} \tag{8}$$

where

$$\omega_0{}^2 = \frac{1}{C_1C_2R_1(R_2 + R_3)} \tag{9}$$

and

$$\frac{\omega_0}{Q} = \frac{1 + \dfrac{R_1}{R_2 + R_3} + \dfrac{C_2}{C_1} - \dfrac{R_3\left[1 + R_f\left(\dfrac{1}{R_A} + \dfrac{1}{R_B}\right)\right]}{k(R_2 + R_3)}}{C_2R_1}.$$

It is convenient to set $1/C_2R_1 = \omega_0$ and $C_2 = C_1$, $R_2 + R_3 = R_1$, whereupon

$$\frac{1}{Q} = 3 - \frac{R_3\left[1 + R_f\left(\dfrac{1}{R_A} + \dfrac{1}{R_B}\right)\right]}{k(R_2 + R_3)}. \tag{10}$$

As a result, a triple staggered-tuned filter may be obtained through use of only three modified dc amplifiers and three cathode followers instead of the nine to twelve dc amplifiers required for synthesizing the filter on an analog computer.

Other Techniques

An alternate procedure for obtaining power spectral measurements involves the recording of data on magnetic tape at slow speed and playing back at significantly faster speeds (Pierson, Jr.[10]). Such a technique has the advantage of speeding the data reduction process and of raising frequencies to a point where standard equipment is available for performing the filtering. The only disadvantage of this technique is the introduction of noise during the process of recording and speeding up the data.

Prediction Theory

The error that results from predicting a random signal $f(t)$ by means of an arbitrary linear time-invariant network with an impulsive response $K_\alpha(t)$ is derived as follows. The normalized mean-squared error of predicting α seconds ahead is given by

[10] W. J. Pierson, Jr., "An Electronic Wave Spectrum Analyzer and Its Use in Engineering Problems," Beach Erosion Board, Corps of Eng., U. S. Army Tech. Memo. No. 56; October, 1954.

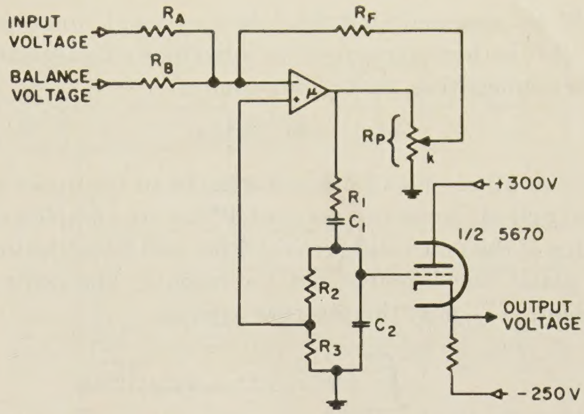


Fig. 7—Band-pass filter (block diagram).

$$E_n(\alpha) = \lim_{T\to\infty} \frac{\dfrac{1}{2T}\displaystyle\int_{-T}^{T}\left[f(t+\alpha) - \int_0^\infty f(t-\tau)K_\alpha(\tau)d\tau\right]^2 dt}{\dfrac{1}{2T}\displaystyle\int_{-T}^{T} f^2(t)dt}. \tag{11}$$

If the transfer function of the prediction network is defined by

$$k_\alpha(\omega) = \int_0^\infty K_\alpha(\tau)e^{-i\omega\tau}d\tau \tag{12}$$

and if the Fourier transform of the truncated signal

$$f_T(t) = \begin{cases} f(t), & -T < t < T \\ 0, & \text{elsewhere} \end{cases} \tag{13}$$

is given by

$$F_T(\omega) = \int_{-T}^{T} f(t)e^{-i\omega t}dt, \tag{14}$$

then by the Parseval relationship, the normalized mean-squared error becomes

$$E_n(\alpha) = \frac{\displaystyle\lim_{T\to\infty}\frac{1}{2T}\int_{-\infty}^{\infty} |F_T(\omega)e^{i\omega\alpha} - F_T(\omega)k_\alpha(\omega)|^2 d\omega}{\displaystyle\lim_{T\to\infty}\frac{1}{2T}\int_{-\infty}^{\infty} |F_T(\omega)|^2 d\omega.} \tag{15}$$

But[9] the power spectrum of $f(t)$ is given by

$$\Phi(\omega) = \lim_{T\to\infty}\frac{|F_T(\omega)|^2}{2T} \tag{16}$$

which reduces the error expression to

$$E_n(\alpha) = \frac{\displaystyle\int_{-\infty}^{\infty}\Phi(\omega)\,|e^{i\omega\alpha} - k_\alpha(\omega)|^2 d\omega}{\displaystyle\int_{-\infty}^{\infty}\Phi(\omega)d\omega}. \tag{17}$$

It will be assumed that $\Phi(\omega)$ is a rational function in ω^2, and it is then convenient to introduce a factorization of the nonnegative, "even" function

$$\Phi(\omega) = \Psi(\omega)\Psi^*(\omega) \tag{18}$$

such that all zeros and poles of $\Psi(\omega)$ lie in the upper half plane, and all zeros and poles of $\Psi^*(\omega)$ are complex conjugates of the poles and zeros of $\Psi(\omega)$ and lie in the lower half plane (see Wiener[2] and Levinson[4]). The error expression, (17), may then be rewritten as

$$E_n(\alpha) = \frac{\int_{-\infty}^{\infty} \left| \Psi(\omega)\left[e^{i\omega\alpha} - k_\alpha(\omega)\right] \right|^2 d\omega}{\int_{-\infty}^{\infty} \left| \Psi(\omega) \right|^2 d\omega}. \tag{19}$$

Now $\psi(t)$, the impulsive response function or weighting function corresponding to the transfer function $\Psi(\omega)$, may be obtained from

$$\psi(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Psi(\omega)e^{i\omega t}d\omega. \tag{20}$$

It is readily shown that $\psi(t)$ vanishes for negative t, since $\Psi(\omega)$ is free of poles in the lower half plane. Thus, by a second application of the Parseval relationship,

$$E_n(\alpha) = \frac{\int_{-\infty}^{\infty}\left[\psi(t+\alpha) - \int_0^{\infty}\psi(t-\tau)K_\alpha(\tau)d\tau\right]^2 dt}{\int_0^{\infty}\psi^2(t)dt} \tag{21}$$

which may be rewritten as

$$E_n(\alpha) = \frac{\int_{-\alpha}^{0}\psi^2(t+\alpha)dt}{\int_0^{\infty}\psi^2(t)dt} + \frac{\int_0^{\infty}\left[\psi(t+\alpha) - \int_0^{\infty}\psi(t-\tau)K_\alpha(\tau)d\tau\right]^2 dt}{\int_0^{\infty}\psi^2(t)dt}. \tag{22}$$

$$\Psi(\omega) = \frac{0.0195(-\omega^2 + 0.820j\omega + 1.611)(-\omega^2 + 0.520j\omega + 0.089)}{(j\omega + 0.616)(-\omega^2 + 0.123j\omega + 0.302)(-\omega^2 + 0.155j\omega + 0.476)} \tag{25}$$

It will be noted that the two error components in (22) are each nonnegative. The first fraction is completely specified by the power spectrum. The second fraction may be reduced to zero, yielding Wiener's optimum predictor formula

$$k_\alpha(\omega) = \frac{1}{\Psi(\omega)}\int_0^{\infty}\psi(t+\alpha)e^{-i\omega t}dt. \tag{23}$$

The first fraction is therefore the inherent least mean-squared predictor error. However, if a nonoptimum weighting function, $K_\alpha(\tau)$, is selected, the second fraction indicates the incremental error that results from this nonoptimum selection.

Verification of the Predictor Design

The formulas that characterize the transfer function of the predictor require an analytic approximation to the power spectral measurement. This approximation is most conveniently performed after the measured data have been plotted on log-log paper. By definition, power spectra are nonnegative even functions of the frequency ω. It is most useful to approximate the power spectrum by a rational function of ω^2, and it is necessary to assume that there are no real poles (corresponding to purely periodic components) nor real zeros (corresponding to frequencies that are absent). In addition, the highest power of ω^2 in the denominator should exceed the highest power of ω^2 in the numerator. Approximation of the power spectrum by a rational function utilizes techniques that are frequently employed in servomechanism problems. Accuracy of the approximations is limited only by the permissible complexity of the analytic approximation.

In the present example, the measured power spectrum (Fig. 6) was approximated by the expression:

$$\Phi(\omega) = \frac{(0.197\omega)^4}{(\omega^2 - 0.37905)^4 + (0.197\omega)^4} + \frac{0.00379}{\omega^2 + 0.37905}, \tag{24}$$

(see Fig. 8).

It is not difficult to obtain

from this power spectrum. Direct application of Wiener's formula, (23), yields the transfer function $k_5(\omega)$ for an optimum five-second predictor, where

$$k_5(\omega) = \frac{1.259(j\omega)^4 + 2.787(j\omega)^3 + 2.553(j\omega)^2 + 1.395j\omega + 0.350}{(j\omega)^4 + 1.340(j\omega)^3 + 2.127(j\omega)^2 + 0.910j\omega + 0.144}. \tag{26}$$

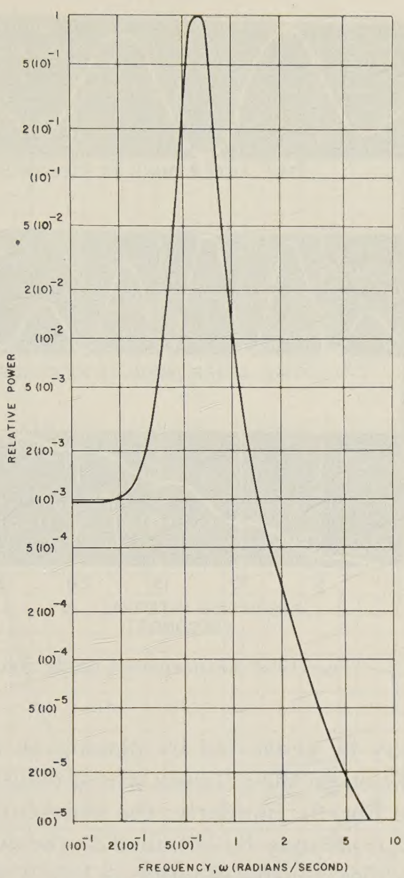
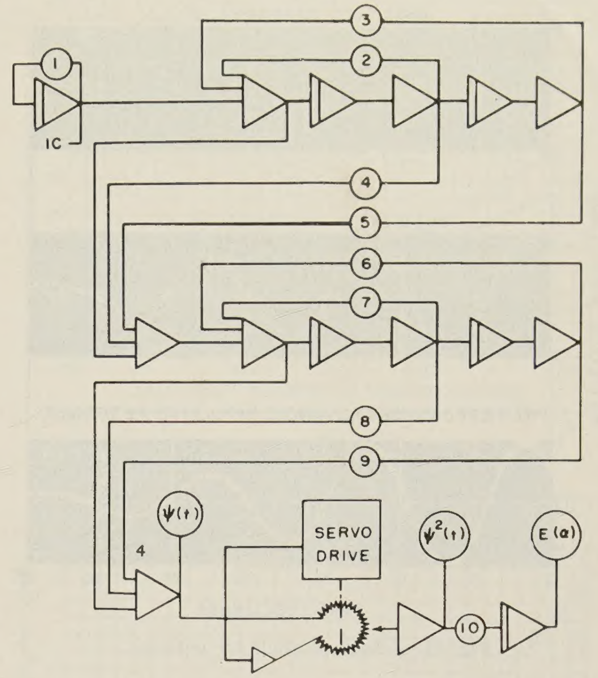


Fig. 8—Power spectrum, analytic approximation.

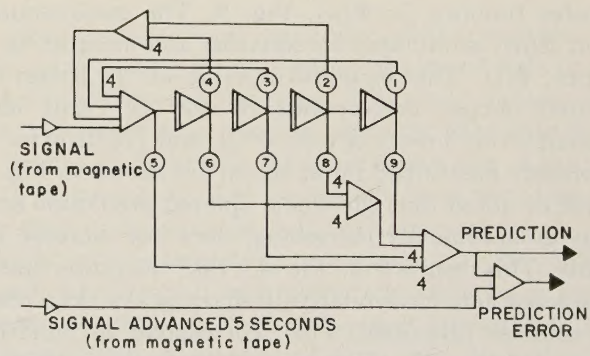


POTENTIOMETER	1	2	3	4	5
SETTING	0.616	0.123	0.302	0.520	0.089

POTENTIOMETER	6	7	8	9	
SETTING	0.476	0.155	0.820	0.403	

SET POTENTIOMETER 10 FOR FULL SCALE DEFLECTION OF E (α) RECORDER AS $\alpha \longrightarrow \infty$

Fig. 9—Circuit for theoretical error.



POTENTIOMETER	1	2	3	4	5
SETTING	0.144	0.910	0.532	0.335	0.315

POTENTIOMETER	6	7	8	9	
SETTING	0.697	0.638	0.349	0.350	

Fig. 10—Wiener predictor network.

The selected predictor transfer function may be verified by comparing $\psi(t+\alpha)$ with the convolution integral $\int_0^\infty \psi(t-\tau)K_\alpha(\tau)d\tau$ [see (22)]. This check may be performed in the following manner. Wire a network with a transfer function $\Psi(\omega)$ on the analog computer (Fig. 9) and apply an impulse or appropriate initial condition to this network to obtain $\psi(t)$. Supply $\psi(t)$ to a second analog computer network corresponding to the predictor (Fig. 10). The resulting predictor response should be an *exact* α-second extrapolation of $\psi(t)$. This check was performed for the five-second predictor (26) corresponding to the power spectral approximation of (24). The comparison of $\psi(t)$, $\psi(t+5)$, and the predictor response to $\psi(t)$ is shown in Fig. 11, next page. It should be noted that this technique may also be justified by using the point of view developed by Bode and Shannon.[3]

Calculation of Inherent Mean-Squared Error

The normalized mean-squared error associated with the optimum linear α-second predictor of (22) may be written as

$$E_n(\alpha) = \frac{\int_0^\alpha \psi^2(t)dt}{\int_0^\infty \psi^2(t)dt} \tag{27}$$

where $\psi(t)$ is the impulsive response of $\Psi(\omega)$, as defined earlier. It is assumed that no purely periodic components are present; $\psi(t)$ decays to zero and therefore, the asymptotic value of $E_n(\alpha)$ is unity, corresponding to 100 per cent error. It should be noted that this process is, in a sense, a trivial case of the adjoint method described by Laning and Battin.[6,7] However, since the present

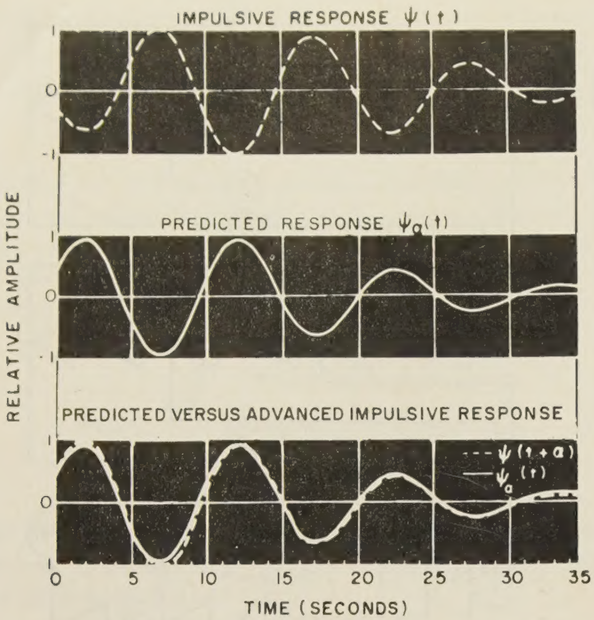


Fig. 11—Check of predictor network.

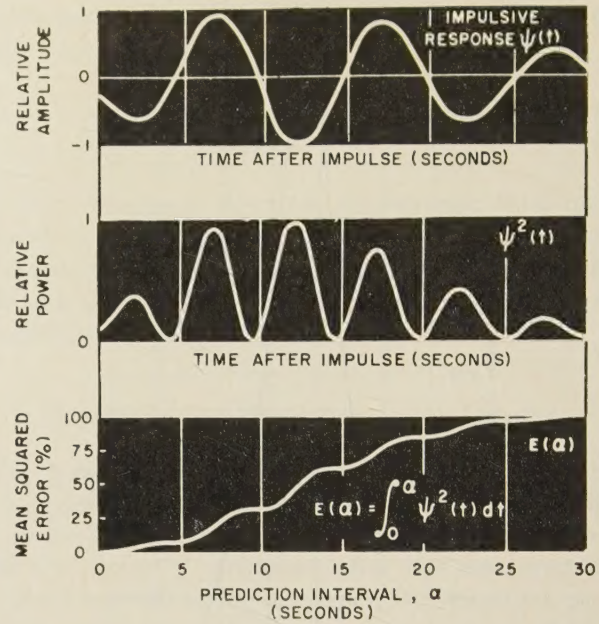


Fig. 12—Theoretical mean-squared prediction error.

paper deals with time-invariant systems, it is possible to justify the method without employing the powerful tools of the adjoint technique. The impulse response method described here can be readily understood on its own merits, as a direct consequence of (22).

Application of this formula for analog computation of the mean-squared error presents no difficulties. The impulsive response, $\psi(t)$, may be generated by imposing an appropriate initial condition on the network whose transfer function is $\Psi(\omega)$, Fig. 9. The mean-squared error, $E(\alpha)$, is obtained by squaring and integrating the output, $\psi(t)$. The asymptotic value of the integrated squared output corresponds to 100 per cent mean squared error. Traces of $\psi(t)$, $\psi^2(t)$, and $\int_0^\alpha \psi^2(t)dt$ for the previously mentioned input signal are shown in Fig. 12. It will be noted that the mean-squared prediction error, while monotonically increasing, does not increase uniformly. This results from the fact that the power spectra considered here have relatively sharp peaks. As a result, the transfer functions, $\Psi(\omega)$, correspond to band-pass filters, and impulsive responses, $\psi(t)$, appear in the form of decaying oscillations. The squaring process doubles the apparent frequency and accentuates the zero crossings. Therefore, the error function $E(\alpha)$, corresponding to the integral of $\psi^2(t)$, grows in a stepwise manner. Thus, in the present example, the error to be expected from predicting 13 seconds ahead is considerably larger than the anticipated error that will result from predicting 11 seconds ahead, whereas the error to be expected in predicting 15 seconds ahead is hardly larger than the error that will result from predicting 13 seconds ahead.

Calculation of the Incremental Error

In order to calculate the incremental error that results from using a nonoptimum predictor, it is again necessary to obtain the impulsive response $\psi(t)$. This response may be generated by placing an appropriate initial condition on the network whose transfer function is $\Psi(\omega)$ (see Fig. 9). Similarly, the convolution integral $\int_0^\infty \psi(t-\tau)K_\alpha(\tau)d\tau$ may be obtained as the output of the predictor [whose transfer function is $k_\alpha(\omega)$] when excited by $\psi(t)$. If $\psi(t)$ and $\psi(t+\alpha)$ are generated simultaneously, the inherent error for a particular prediction interval, and the incremental error

$$\frac{\int_0^\infty \left[\psi(t+\alpha) - \int_0^\infty \psi(t-\tau)K_\alpha(\tau)d\tau\right]^2 dt}{\int_0^\infty \psi^2(t)dt}$$

resulting from use of a nonoptimum predictor may be readily calculated on an analog computer through the use of subtracting, squaring, and integrating circuits. Simultaneous generation of $\psi(t)$ and $\psi(t+\alpha)$ is readily achieved by wiring identical networks $\Psi(\omega)$ on each of two computers, applying the same initial conditions, and employing a relay to turn one machine on exactly α seconds after the other machine goes on.

A specific power spectrum of the form

$$\Phi(\omega) = \frac{c^2\omega^6 + B^2\omega^4 + A^2}{(\omega^2 - \omega_0^2)^4 + \delta^4\omega^4}$$

$$= \frac{0.3791\omega^6 + 1.2188\omega^4 + 0.0206}{(\omega^2 - 0.3791)^4 + 0.001506\omega^4}, \tag{28}$$

was used for selecting the predictor, and the incremental error was determined for three groups of perturbed power spectra. In the first perturbation, ω was replaced by $\omega/(1+\epsilon)$, and ϵ was varied from $+0.2$ to -0.2. In the second perturbation, δ was varied between 0.1 and 0.6

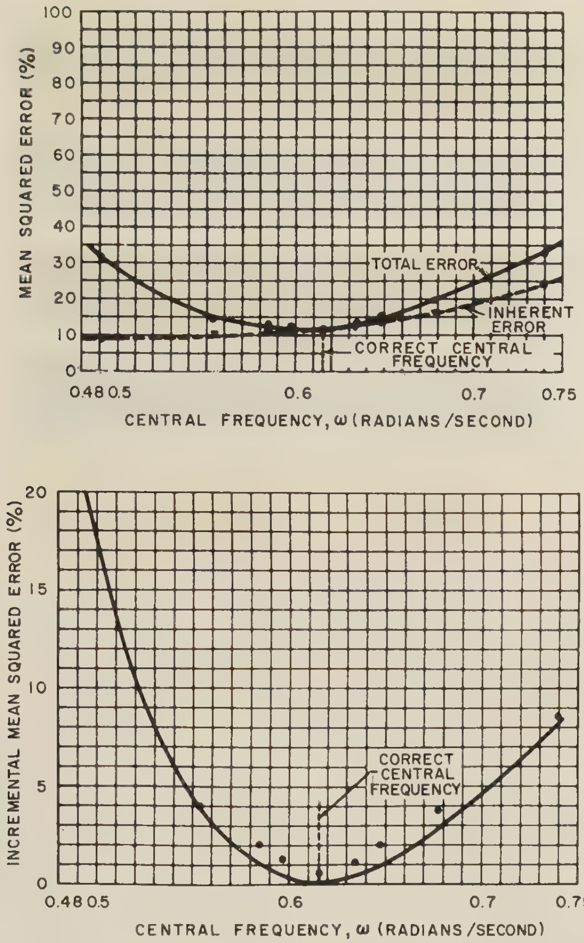


Fig. 13—Error increment from wrong central frequency assumption.

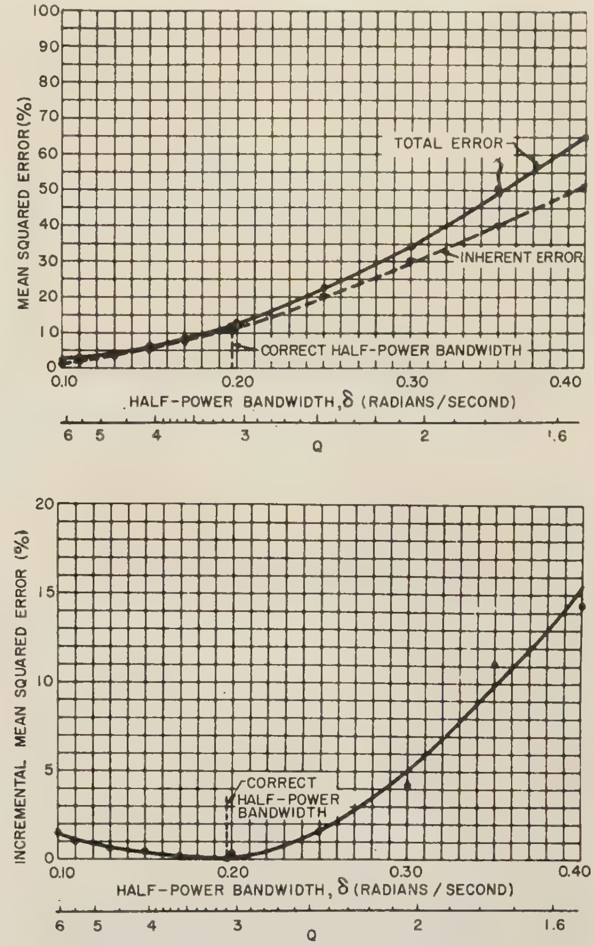


Fig. 14—Error increment from wrong bandwidth assumption.

(design bandwidth = 0.197). In the third perturbation, A^2 was maintained equal to $C^2\omega_0^6$, but the ratio A^2/B^2 (arbitrarily called the dc level) was varied from 0.00001 to 0.005 (design value = 0.001). In Fig. 13, Fig. 14, and Fig. 15, next page, the inherent error corresponding to each of these assumed power spectra is compared with the total error by utilizing the preselected predictor. The inherent errors varied considerably, but the incremental errors were significant only when the central frequency changed by more than 5 to 10 per cent when the half-power bandwidth exceeded the design value by more than 30 to 50 per cent, or when the low and high-frequency power varied excessively.

In view of these results, it appears that an accurate central frequency measurement is essential, and that rough (average) estimates of half-power bandwidth, dc signal content, and high-frequency cutoff may be adequate. An investigation of the effects of small discrepancies in the magnitudes of selected passive elements of the predictor may be performed in a similar manner.

By using a servo to set a series of ganged pots to the measured central frequency ω_0, it has been proven possible to design a time varying predictor that adjusts itself to slow variations in the power spectrum of the signal to be predicted. The device has been tested in the laboratory and also in the field with some success.

Similar error-calculating techniques may be applied to predicting a signal with power spectrum $\Phi_s(\omega)$ in the presence of additive noise with power spectrum $\Phi_n(\omega)$. The normalized mean-squared error may be written as

$$E_n(\alpha) = \frac{\int_{-\infty}^{\infty} \left[\left| e^{i\omega\alpha} - k_\alpha(\omega) \right|^2 \Phi_s(\omega) + \left| k_\alpha(\omega) \right|^2 \Phi_n(\omega) \right] d\omega}{\int_{-\infty}^{\infty} \Phi_s(\omega) d\omega} . \quad (29)$$

Application of the Parseval relationship yields

$$E_n(\alpha) = \frac{\int_{-\alpha}^{0} \psi_s^2(t+\alpha)dt + \int_{0}^{\infty} [\psi_s(t+\alpha) - K_\alpha(t) * \psi_s(t)]^2 dt + \int_{0}^{\infty} [K_\alpha(t) * \psi_n(t)]^2 dt}{\int_{0}^{\infty} \psi_s^2(t) dt} \quad (30)$$

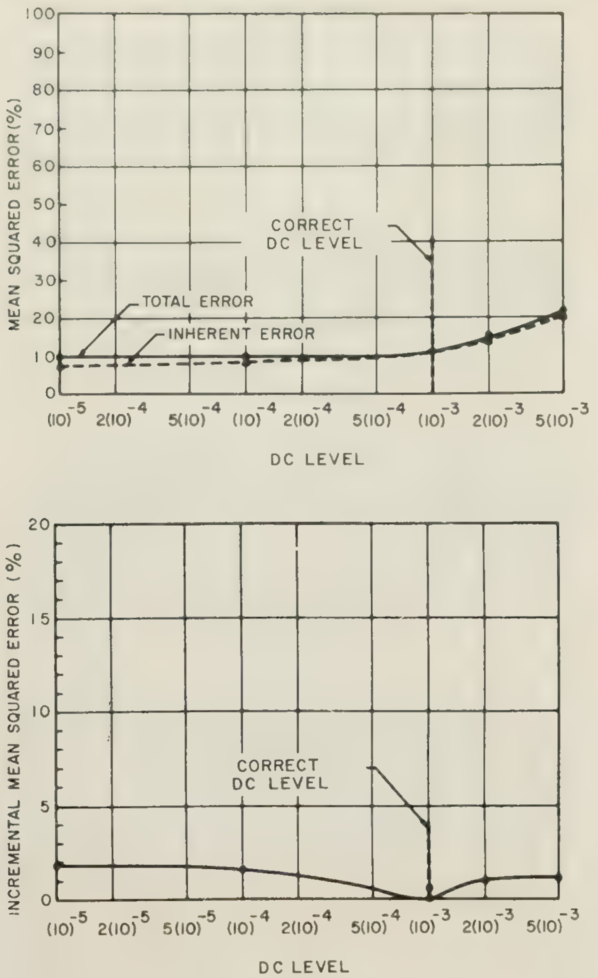


Fig. 15—Error increment from wrong assumption or dc level.

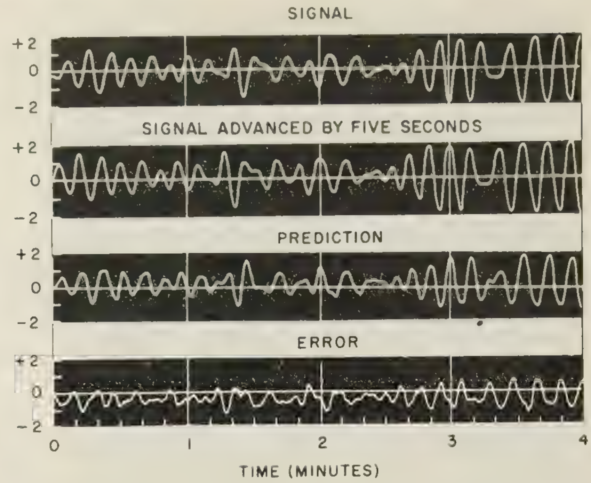


Fig. 16—Predictor performance.

where the convolution operation $\int_0^\infty \psi(t-\tau)K_\alpha(\tau)d\tau$ is designated by $K_\alpha(t)*\psi(t)$. This error term can also be calculated on an analog computer, using the impulsive response techniques mentioned earlier.

Final Check

The ultimate check on the accuracy of the predictor is obtained by an actual prediction of the recorded signal. The recorded signal is supplied as an input to the predictor, and the predictor output is stored or delayed for a time corresponding to the prediction interval, α, and is compared with the input signal α seconds in the future.

An exact measurement of the prediction error may thus be obtained by utilizing a delay line. The magnetic tape passes over two readout heads that are separated by a distance that corresponds to α seconds at the appropriate tape speed. The head furthest from the take-up reel supplies advance information for comparison with the predictor output. The instantaneous prediction error may be obtained by means of an appropriate differencing circuit. This error is shown together with the input and output of the predictor and an exact α-second extrapolation of the predictor input (Fig. 16). Finally, the mean-squared prediction error may be measured by conventional analog computer techniques and compared with the theoretically determined mean-squared error for the optimum predictor. In this case, the measured mean-squared error was about 11.5 per cent whereas the theoretically calculated mean-squared error was about 11 per cent which seems to indicate excellent agreement.

Once the various checks have been run, the predictor is available for use in the form of a circuit on the analog computer. A network is then available for predicting any signal that has the measured power spectrum. Thus, if the data are statistically stationary, samples thereof may be analyzed, and the corresponding predictor network may be designed and utilized to predict other samples of the same data. If it is desirable, some dc amplifiers may be saved at the expense of additional network complexity.

Appendix

Additional Results of Error Computations

The analog computation of mean-squared error may also be utilized to indicate the theoretical dependence of mean-squared error on such power spectral parameters as central frequency, half-power bandwidth, low-frequency content, and high-frequency cutoff. The dependence of predictor error on central frequency and half-power bandwidth is illustrated by considering the family of band-pass spectra

$$\Phi(\omega) = \frac{\omega^2/\omega_0^2Q^2}{(1-\omega^2/\omega_0^2)^2 + \omega^2/\omega_0^2Q^2} \tag{31}$$

discussed by Lee and Stutt.[11] These spectra are shown in Fig. 17 and are characterized by central frequency ω_0 and ratio of central frequency to half-power bandwidth,

[11] Y. W. Lee and C. A. Stutt, "Statistical Prediction of Noise," M.I.T. Res. Lab. Electronics Rep. No. 129; 1949.

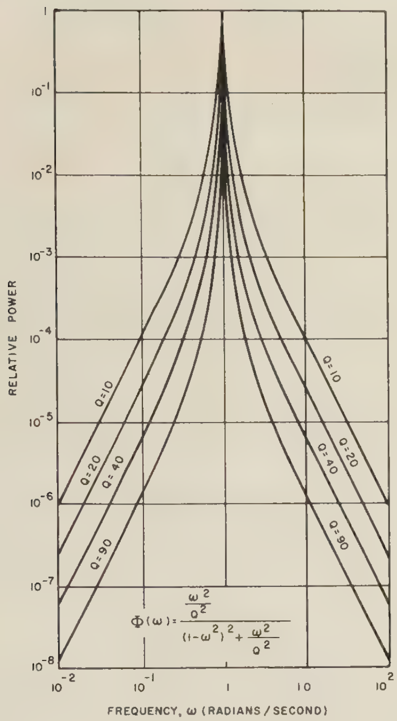


Fig. 17—Band-pass power spectra.

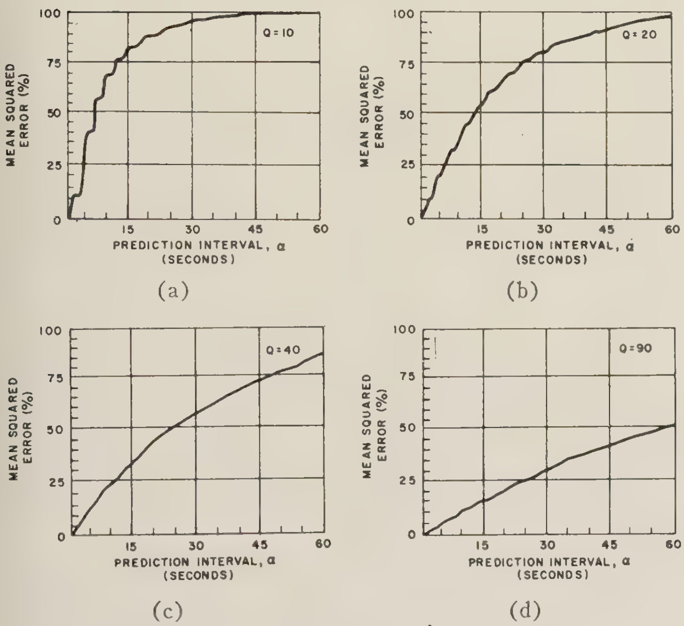


Fig. 18—Mean-squared error for band-pass power spectra.

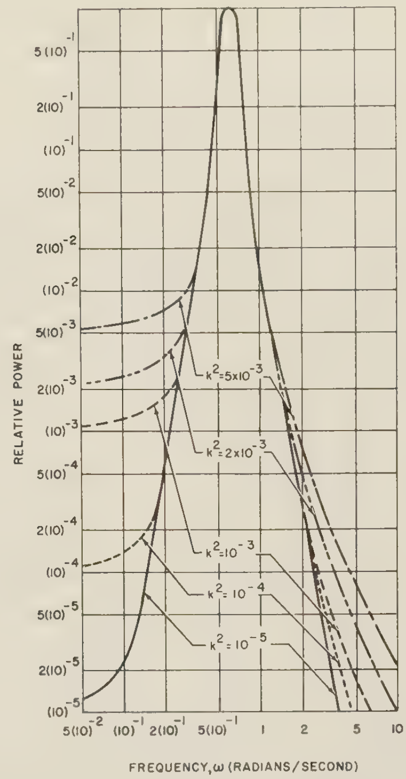


Fig. 19—Family of power spectral approximations.

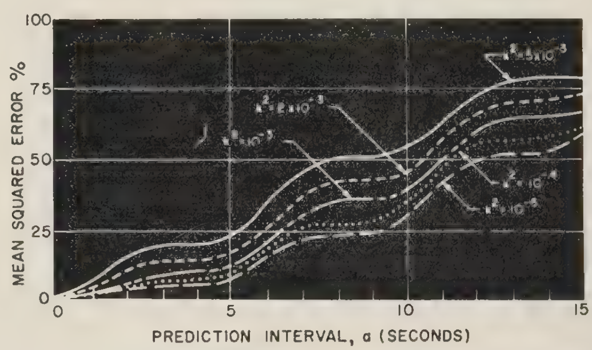


Fig. 20—Mean-squared error for various spectral approximations.

Q. The power spectra are functions of ω/ω_0, and consequently, the mean-squared error is a function of $\omega_0 t$. A plot of mean-squared error for Q equal to 10, 20, 40, and 90 is shown as Fig. 18.

The dependence of prediction error on dc content and high-frequency cutoff rate is illustrated by consideration of the family of power spectra

$$\Phi(k, \omega) = \frac{k^2(\omega_0^8 + \omega^6\omega_0^2) + (\delta^4 - 2k^2\omega_0^4)\omega^4}{(\omega^2 - \omega_0^2)^4 + \delta^4\omega^4} \quad (32)$$

where fixed values of δ^2 and ω_0^2 were used. The ratio of dc power to peak power is approximately k^2 (Fig. 19). Corresponding mean-squared error traces were obtained on the analog computer and are shown in Fig. 20.

The relationship between prediction accuracy and power spectral parameters corresponds in its general nature to the results that would be anticipated from physical reasoning. Thus, in dimensionless time, $\omega_0 t$, accuracy of prediction is strongly dependent on the sharpness of the spectrum, improving rapidly with increasing Q, decreasing low-frequency signal content, and sharper high-frequency cutoff. The numerical results presented in Figs. 18 and 20 indicate the precise nature of this dependence. Similar computations may be necessary to determine corresponding results for other spectra of specific interest.

The Theory of Nets*

F. E. HOHN†, S. SESHU‡, AND D. D. AUFENKAMP§

Summary—This paper presents the general concept of a weighted directed graph which we call a net. Illustrative examples leading up to the definition of a net indicate its applicability to a wide variety of problems in communication and network theory. A number of theorems concerning the proper paths of a net are established. A non-arithmetic matrix calculus is developed to facilitate computations and formalize proofs. In later papers, the techniques presented here will be exploited in the study of the theory of sequential machines.

I. Illustrative Examples

A. The Connection Relation of a Finite Linear Graph

The origin of one theory of sequential machines is ultimately in the theory of finite, nondirected, linear graphs which, in accordance with custom, we simply refer to as *linear graphs* in the following discussion. Geometrically, such a graph is a finite set of nodes—points in 3-space—and a finite set of branches connecting certain of these nodes and having no other points in common than the nodes. Branches proceeding from a node back to the same node are customarily excluded from consideration, but in the generalizations to be developed in later sections, they are often useful so we choose to admit them. The fundamental relation between node and branch is that of *incidence.* The information regarding incidence is summarized in what is called an *incidence matrix* in which there is one row for each node and one column for each branch. Suppose the nodes are numbered 1, 2, $\cdots$, n and the branches 1, 2, $\cdots$, m. The elements of the incidence matrix H_1 are then defined as follows:

$$H_1 = [h_{ij}]_{nxm}, \; h_{ij} = \begin{cases} 1 \text{ if node } i \text{ and branch } j \text{ are incident,} \\ 0 \text{ otherwise.} \end{cases}$$

From this incidence matrix one derives many of the essential properties of linear graphs [1, 2].

In the matrix product,

$$R_1 = H_1 H_1^T,$$

formed in the usual way, we have as the entry in the ith row and the jth column

$$r_{ij} = \sum_{k=1}^{m} h_{ik} h_{jk}.$$

The term $h_{ik}h_{jk}$ of this sum will be 1 if and only if $h_{ik} = h_{jk} = 1$; that is, if and only if nodes i and j are both on branch k. If $i \neq j$, $h_{ik}h_{jk}$ is thus 1 if and only if nodes i and j are *connected* by branch k. Thus if ordinary arithmetic is employed, r_{ij} is equal to the number of branches connecting i and j, when $i \neq j$. On the other hand, when $i = j$, r_{ij} is equal to the number of branches incident with node i. Thus if there are no branches from node i to itself, $r_{ii} > 0$ if and only if node i is not isolated from the others.

If we interpret "+" as the Boolean "or" so that $0+0=0$, $0+1=1+0=1+1=1$, $0\cdot 0=0\cdot 1=1\cdot 0=0$, $1\cdot 1=1$, in evaluating the entries of R_1, we have

$$r_{ij} = \begin{cases} 1 \text{ if there exists at least one branch} \\ \quad \text{connecting nodes } i \text{ and } j \text{ when } i \neq j; \\ 1 \text{ if } i \text{ is not isolated or has a loop, when } i = j, \\ 0 \text{ otherwise.} \end{cases}$$

The matrix R_1 is then the "relation matrix" for the relation of connectedness of the nodes of the graph. Since in a nondirected graph the connection relation is symmetric, R_1 will also be symmetric. We call R_1 briefly the *connection matrix* of the graph.

B. Binary Relations on a Finite Set

The connection matrix of a finite linear graph suggests a method of representing any binary relation on a finite set. A binary relation on a finite set S of order n is defined to be a subset of the set of all ordered pairs of elements of S. The pair (s_i, s_j) belongs to the subset; *i.e.*, to the relation, if and only if s_i bears the relation to s_j. We can represent the set S as the nodes of a graph in which each ordered pair (s_i, s_j) of the relation is represented by a *directed* branch *from* the node corresponding to s_i *to* the node corresponding to s_j. In this graph there is at most one directed branch from one node to another. We can now describe both the relation and the associated graph by means of a relation matrix R of order n:

$$R = [r_{ij}]_n, \; r_{ij} = \begin{cases} 1 \text{ if } (s_i, s_j) \text{ belongs to the relation,} \\ 0 \text{ otherwise.} \end{cases}$$

Such relation matrices have been studied extensively [3, 4].

In essence, a relation matrix R defines all the properties of the corresponding relation. For example, a relation is *symmetric* if and only if for each pair (s_i, s_j) belonging to it, the pair (s_j, s_i) also belongs to the relation. Thus the relation is symmetric if and only if its relation matrix R is symmetric. Again, a relation is *reflexive* if and only if every pair (s_i, s_i) belongs to the relation; *i.e.*, if and only if $r_{ii} = 1$ for all i. In the graph of a relation, to a pair (s_i, s_i) there corresponds a *loop*; *i.e.*, a branch proceeding from the node corresponding to s_i back to the same node. A relation is *transitive* if and only if

* Manuscript received by the PGCT, December 18, 1956; revised manuscript received, June 19, 1957.

† University of Illinois, Urbana, Illinois.
‡ Syracuse University, Syracuse, N. Y. Formerly with the University of Illinois, Urbana, Ill.
§ Lockheed Missile Systems Division., Sunnyvale, Calif.

whenever (s_i, s_j) and (s_j, s_k) belong to the relation, (s_i, s_k) belongs to the relation also. Now we define the symbol "$\leq$" for "inclusion" in Boolean arithmetic to mean $0 \leq 0$, $0 \leq 1$, $1 \leq 1$, and define further for two matrices R and P of the same order that $R \leq P$ if and only if $r_{ij} \leq p_{ij}$ for all i and j. Then it is not hard to show that a relation with matrix R is transitive if and only if $R^2 \leq R$ where R^2 is formed in the usual way but with Boolean arithmetic being employed. Finally, using Boolean arithmetic and the logical product $R*S = [r_{ij}s_{ij}]$ for matrix multiplication, we could show that the set of all relation matrices of order n is a Boolean algebra with 2^{n^2} elements ([4], p. 209). The purpose of these remarks is simply to emphasize that the properties of a relation may be deduced from suitable operations on its matrix R.

The concept of a finite relation underlies much of the work of the papers of this sequence.

C. *Combinational Relay Circuits*

Let us next look at a combinational relay-contact switching circuit (for details see [5, 6]) such as is illustrated in Fig. 1, in which a, $\bar{b}$, c, x denote contacts. Here again we have an underlying graph, certain nodes

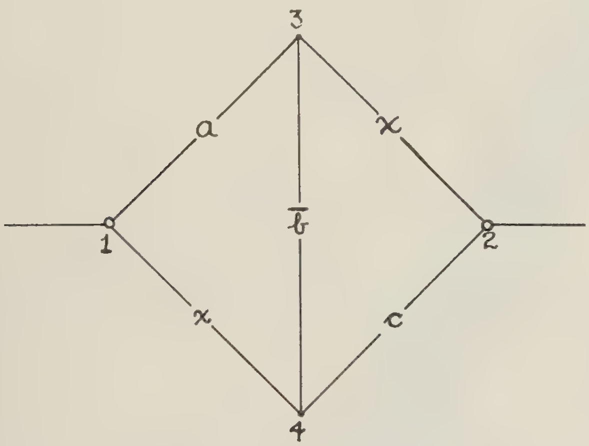


Fig. 1.

of which—the *terminal nodes*—the network is designed to interconnect, and other nodes of which—the *nonterminal nodes*—merely represent binding posts in the circuitry. In Fig. 1, nodes 1 and 2 are terminal nodes, whereas nodes 3 and 4 are nonterminal.

Now assume that exactly enough nonterminal nodes have been marked on the graph so that between any two nodes, terminal or nonterminal, there are no two contacts in series; *i.e.*, at most a single contact or contacts in parallel. All these nodes together constitute a finite set and the condition that certain pairs of them be connected when a given contact is closed is a relation on this set, associated with that particular contact symbol. That is, if there are p nodes altogether, then for each contact symbol x of the switching circuit, there exists a $p \times p$ relation matrix

$$R^x = [r_{ij}{}^x]$$

such that

$$r_{ij}{}^x = \begin{cases} 1 & \text{if } i \neq j \text{ and } i \text{ and } j \text{ are connected when } x \text{ is closed,} \\ 1 & \text{if } i = j \text{ since in this application a node is assumed to be always connected to itself,} \\ 0 & \text{otherwise.} \end{cases}$$

(If contacts $\bar{x}$ appear, where $\bar{x}$ is the complement of x, then a matrix $R^{\bar{x}}$ is associated in the same way with $\bar{x}$.) For example, in the case of the circuit of Fig. 1, we have

$$R^x = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

Since relays are bilateral devices, these relation matrices are always symmetric.

The *primitive connection matrix* of the circuit may now be defined as a weighted sum of these relation matrices. If x_j, $\bar{x}_j$, $(j = 1, 2, \cdots, n)$ are the contact variables, this sum is defined to be

$$P = \sum_{j=1}^{n} x_j R^{x_j} + \sum_{j=1}^{n} \bar{x}_j R^{\bar{x}_j}$$

where the symbol "+" is the "inclusive or" operation of Boolean algebra. (The ij entry of xR^x is defined to be xr_{ij}, if $i \neq j$ and 1 if $i = j$.)

In the case of our example, we have

$$P = aR^a + \bar{b}R^{\bar{b}} + cR^c + xR^x = \begin{bmatrix} 1 & 0 & a & x \\ 0 & 1 & x & c \\ a & x & 1 & \bar{b} \\ x & c & \bar{b} & 1 \end{bmatrix}$$

which can be determined directly from the circuit by inspection.

The algebra governing the contact symbols is the usual Boolean algebra of switching circuits. The primitive connection matrix is in effect a wiring diagram of the circuit.

A primitive connection matrix is an example of what is called a *switching matrix*, namely, a square matrix whose diagonal entries are all 1's and whose off-diagonal entries are arbitrary switching functions. These switching matrices may be employed effectively in both the analysis and the synthesis of combinational relay circuits [6, 7].

II. Sequential Machines

A. *State Diagrams of Sequential Machines*

The same concept of a connection relation enters into the theory of *sequential machines*; that is, devices or circuits which are designed to make certain sequences of outputs correspond to given sequences of inputs. A simple, abstract mathematical model of such a machine has recently been proposed by Moore of Bell Telephone

Laboratories [8]. In Moore's model, a sequential machine is defined as consisting of

a finite number of distinct *states*, $s_1, s_2, \cdots, s_n$,
a finite number of distinct *inputs*, $x_1, x_2, \cdots, x_p$, and
a finite number of distinct *outputs*, $y_1, y_2, \cdots, y_q$,

all related in such a way that 1) the present output depends only on the present state and 2) the present state depends only on the previous state and the previous input. That is, each input applied to the machine in a given state results in a uniquely defined next state with its uniquely defined output. This definition makes it clear that the present state and output depend on the past history of the machine. In this definition, the word "state" is to be regarded as an undefined term.

Moore represents such a machine by a graph as follows: to each state s_i corresponds a node i; to each distinct input which drives the machine from state s_i to state s_j, there corresponds a directed branch from node i to node j. This graph is called a *state diagram* of the machine. Such a graph may have more than one branch from one node to another and it may also have loops from a node to itself. In drawing it, we denote a node by a circle large enough to include a symbol identifying the node and a symbol identifying the output associated with the corresponding state. Beside each branch is written the symbol for the corresponding input. These matters are illustrated by the state diagram of a ternary flip-flop shown in Fig. 2. Here there are three states 1, 2,

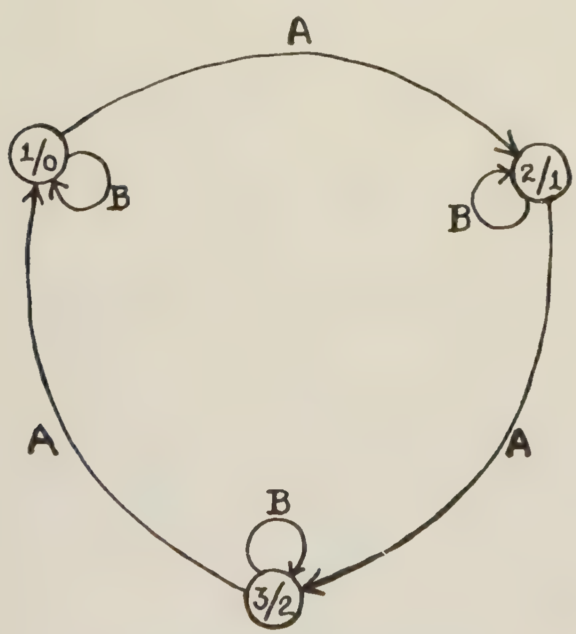


Fig. 2.

3, with associated outputs 0, 1, 2, respectively. The inputs are A and B; an input B leaves each state invariant while the input A drives each state into the "next" state: 1 into 2, 2 into 3, and 3 into 1.

Using the above definition and representation of sequential machines, Moore derives a number of their important properties. In this and succeeding papers, we introduce a matrix calculus which enables us to mechanize and to extend many of his results.

The development of the matrix model here is closely analogous to that given for relay circuitry in the last section. Thus, in the case of our ternary flip-flop, we associate with the inputs A and B the following relation matrices:

$$T^A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \qquad T^B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We call these "transition matrices" because they represent the transitions from one state to another occasioned by the corresponding input. Also, as before, we define a weighted sum of these relation matrices which we call the "connection matrix" C of the machine. In the case of our example, we have

$$C = AT^A + BT^B = \begin{bmatrix} B & A & 0 \\ 0 & B & A \\ A & 0 & B \end{bmatrix}$$

which reveals how the states are permuted by the various inputs.

In general, if the inputs of a machine with n states are $x_1, x_2, \cdots, x_p$, we associate with each x_k the *transition matrix* T^k defined as follows:

$$T^k = [t_{ij}{}^k]_{nxn} \qquad t_{ij}{}^k = \begin{cases} 1 & \text{if input } x_k \text{ takes state } s_i \\ & \text{into state } s_j, \\ 0 & \text{otherwise.} \end{cases}$$

The *connection matrix* of the machine is then defined as

$$C = \sum_{k=1}^{n} x_k T^k.$$

In the formation of this sum, one uses the usual rule for multiplying a matrix by a scalar and the operation of addition has the meaning "or."

It is important to note that giving the outputs of the n states and either the set of transition matrices or the single matrix C defines the machine completely.

In later papers of this sequence, we employ the transition matrices as well as other appropriate matrices to study the analysis and synthesis of sequential machines in detail. In the remainder of this paper, we discuss a generalized type of graph of which each type of graph mentioned so far is a special case.

III. Nets

A. *Definition of a Weighted, Directed Graph or Net*

The concept of a state diagram suggests the following generalization. We begin with a finite set of objects, denoted by $v_1, v_2, \cdots, v_n$ which we call *vertices* and a finite or infinite, discrete or continuous set $\{w\}$ of *vertex*

weights. From $\{w\}$ we select a single element w_i to be associated with each vertex v_i. The set of pairs (v_i, w_i) is the set of *weighted vertices*. The w_i need not all be distinct so when convenience requires, we denote the distinct elements among them by $\omega_1, \omega_2, \cdots, \omega_p$.

We next consider a set of not necessarily distinct ordered pairs of vertices (v_i, v_j), which are called *directed branches*. (A branch of the form (v_i, v_i), is called a *loop*.) Let these branches be m in number and let them be denoted by $b_1, b_2, \cdots, b_m$. Now let there be a finite or infinite, discrete or continuous set $\{x\}$ of *branch weights* from which we select a single element x_j for each branch b_j. Then the set of pairs (b_j, x_j) is the set of *weighted branches*. The x_j need not all be distinct, and we denote the distinct elements among them by $\xi_1, \xi_2, \cdots, \xi_q$.

The sets of weighted vertices and weighted branches defined above together constitute what we call a *weighted, directed graph or net*. This abstract definition of a net has a simple, geometrical representation. As in the case of the state diagram, we represent the vertices as small, disjunct circles in the plane. Inside each circ[illegible] we write the name of the vertex and of the associa[illegible] weight. For each branch of the net we draw an a[illegible] from the initial vertex to the terminal vertex [illegible] branch and write the weight of the branch be[illegible] An example of such a graph is shown in Fig. 3.

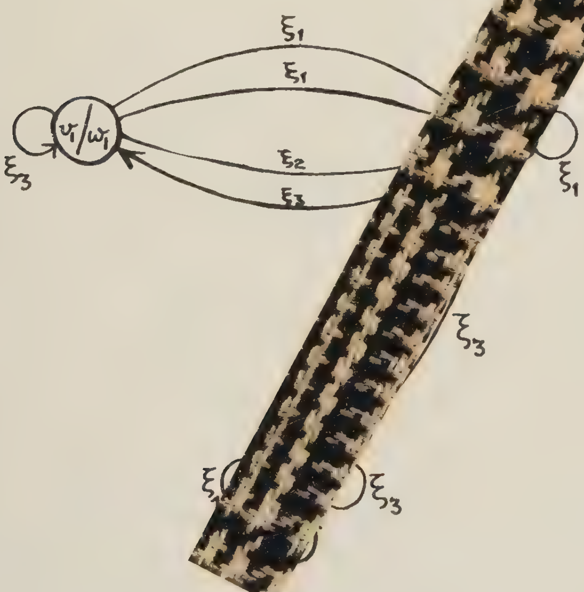


Fig. 3.

If loops do not appear, and if all the vertex weights are the same and all the branch weights are the same, the weights may be ignored and the theory of *directed graphs* ([2], p. 4) results as a special case. If further with each branch (v_i, v_j) the graph also contains the branch (v_j, v_i), we may identify each such pair of oppositely directed branches as a single, undirected branch and thus obtain the usual theory of *linear graphs* as a special case. If no loops appear, if the vertex weights are all equal and hence may be ignored, if the branches appear in equal-weighted pairs so that they may be regarded as undirected, and if the branch weights are switching functions, we have the relay circuit example mentioned in Section I-C as a special case. If the vertex weights are node voltages and the branch weights are currents or impedances, we obtain ordinary network theory. The state diagram of a sequential machine provides a totally different example of a net in which both vertex weights and branch weights are of significance.

It is important to note that each application requires its own special algebra for combining the weights. In the relay case, it is the usual Boolean algebra of switching functions; in the case of ordinary network theory, it is the algebra of complex numbers, etc. Also, it is possible that in particular cases there are different aspects of the physical system under study which result in there being more than one set of branch weights (or node weights). In such a case we may associate more than one net with the problem.

[illegible]rious special nets have been applied to nonelec-[illegible] problems. Some of these concern communication [illegible]blems [9–13]; sociology and psychology [14, 15]; [illegible]emistry [16]; logic [17]; and others. The work of [illegible]himbel [9] has provided inspiration for the present definitions. A somewhat more restricted concept of a net than ours has been studied by Luce [20, 21].

B. The Fundamental Matrices of a Net

In the theory of directed graphs as developed by Veblen [1] and König [2], no loops are admitted and an incidence matrix A is employed which is a generalization of that defined in Section I-A. If there are n vertices and m branches, $A = [a_{ij}]_{nxm}$ is defined as follows:

$a_{ij} = 1$ if branch j has vertex i as an *initial* point.
$a_{ij} = -1$ if branch j has vertex i as a *terminal* point.
$a_{ij} = 0$ if branch j does not have vertex i as an end point.

In the applications it is at times necessary to admit loops. In this case certain entries of A would have to be both 1 and -1 so that the above definition of the incidence matrix would no longer be meaningful. Hence we shall define two other matrices, A^+ and A^-, related to the incidence matrix, which are appropriate for the study of the more general case.

The matrix $A_{nxm}{}^+$ is defined thus:

$a_{ij}{}^+ = 1$ if branch j has vertex i for an *initial* point.
$a_{ij}{}^+ = 0$ otherwise.

The matrix $A_{nxm}{}^-$ is defined thus:

$a_{ij}{}^- = 1$ if branch j has vertex i for a *terminal* point.
$a_{ij}{}^- = 0$ otherwise.

In case the graph contains no loops, the incidence matrix is related to A^+ and A^- by the equation

$$A = A^+ - A^- \text{ (no loops)}$$

where the entries of A^+ and A^- are treated as real numbers.

The matrices A^+ and A^- completely describe the structure of the graph except for the weighting of the vertices and branches. It is most convenient to record the vertex weights in the form of an n vector

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ w_n \end{bmatrix},$$

and the branch weights in the form of a diagonal matrix of order m

$$X = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & x_m \end{bmatrix}.$$

Then the matrices A^+, A^-, W, and X describe the net completely.

In certain applications, to sequential machines, for example, the distinct branch weights form a discrete set and there is at most one branch, which we now denote by (i, j, ξ_k), from a vertex i to the vertex j and having a given weight ξ_k. For the remainder of this section, we assume that these circumstances prevail.

Another class of matrices, useful under the hypotheses just stated, relates the matrix X to the matrices A^+ and A^-. These matrices, called relation matrices, structure matrices, or transition matrices in the various applications, are defined as follows. With each distinct branch weight ξ_k which appears among $x_1, x_2, \cdots, x_m$, we associate a relation matrix $R^k = [r_{ij}{}^k]$ of order n:

$$r_{ij}{}^k = 1 \text{ if } (i, j, \xi_k) \text{ is a branch of the net,}$$

$$r_{ij}{}^k = 0 \text{ if } (i, j, \xi_k) \text{ is } \textit{not} \text{ a branch of the net.}$$

Thus R^k describes the distribution of the branches with weight ξ_k in the net. Rephrased, the definition may be stated thus: $r_{ij}{}^k = 1$ if a branch of weight ξ_k has vertices (i, j) and is oriented from i to j, $r_{ij}{}^k = 0$ otherwise. Again we notice that the set of matrices $R^1, R^2, \cdots, R^p$, together with the vector W of the weights of vertices, is a complete description of the net.

Finally, we associate with a net still another matrix which is very useful for a study of the structure of the graph. This is the *connection matrix* $C = [c_{ij}]$ of order n defined as follows:

$$c_{ij} \doteq \sum{}' x_k$$

where x_k is the weight of a branch from vertex i to vertex j, the "summation" $\sum'$ is over all such branches, and $c_{ij} = 0$ if there is no such branch. The meaning of the operation of "addition" used here depends on the particular application. A similar remark applies to "multiplication." However, in all cases we preserve the rules $x + 0 = 0 + x = x$, $0 \cdot x = x \cdot 0 = 0$, and $1 \cdot x = x \cdot 1 = x$.

We note that the matrices C and W give a third complete description of the net. The relations between the matrices $A^+, A^-, R^1, R^2, \cdots, R^p, C$, and X involved in these descriptions are given in the following two theorems.

Theorem 1: $C = A^+X(A^-)^T$.

Here the superscript T denotes the transpose, and matrix multiplication is performed in the usual manner.

We observe first that since X is diagonal, A^+X differs from A^+ only in that the kth column is multiplied by x_k. This is meaningful since the sole nonzero entry in the jth column of A^+ is a 1 in the ith row where branch j is directed *from* vertex i. Now there is a nonzero entry x_k in the ith row of A^+X with a "1" in the corresponding position of the jth column of $(A^-)^T$ for each branch k which is directed *from* vertex i *to* vertex j. Hence the ij-entry of the product $A^+X(A^-)^T$ is precisely $\sum' x_k = c_{ij}$.

Theorem 2: $C = \sum_{k=1}^{p} \xi_k R^k$.

This follows at once from the definitions of C and the R^k's.

C. The Paths of a Net

A primary concern of our later work will be the paths of a net. A *directed path* of length r from vertex v_i to vertex v_j is a set of r branches of the form (v_i, v_k), (v_{k_1}, v_{k_2}), $\cdots$, $(v_{k_{r-1}}, v_j)$ provided such branches exist in the net. If $i = j$, the path is closed and is called a *cycle* of length r. If $v_i, v_{k_1}, \cdots, v_{k_{r-1}}, v_j$ are all distinct vertices, the path is a *proper path* and v_i is its *initial vertex*, v_j its *terminal vertex*. If $v_i \equiv v_j$ but no other two vertices coincide, the path is a *proper cycle*. In particular, a loop is a proper cycle of length 1. A path or cycle which is not proper is *redundant*.

In a net of n vertices, a proper cycle of length n, if it exists, is called a *directed Hamilton line*. Not all nets contain directed Hamilton lines.

It is clear that whenever there is a redundant path from a vertex v_i to a vertex v_j in a net, then there exists also a shorter, proper path from v_i to v_j. The proper paths of a net are of particular interest in the applications. Accordingly, our first effort is directed toward their systematic identification.

For this purpose we let the synbol c_{ij} itself be a weight for the branch from i to j if there is just one such branch. If there is no such branch, we let $c_{ij} = 0$. If there is more than one branch from i to j, we denote these by $c_{ij}{}^{(1)}, c_{ij}{}^{(2)}, \cdots$ and put $c_{ij} = \sum{}^k c_{ij}{}^{(k)}$, where addition has in effect the meaning "or." For this set of branch weights, $C = [c_{ij}]$ is the connection matrix of the net.

First we consider the matrix C^2, which is formed from C in the usual way, multiplication of the entries being associative and distributive with respect to addition,

but *not* commutative. If we now interpret multiplication to mean "and" and addition to mean "or," the entries

$$\sum_{k=1}^{n} c_{ik}c_{kj} \text{ of } C^2$$

list all the paths of length 2 of the net, both proper and redundant. More generally, we have this theorem.

Theorem 3: The entries of C^q list all paths of length q, both proper and redundant, of the net.

The ij entry of C^q is, in fact,

$$\sum_{(k)} c_{ik_1}c_{k_1k_2} \cdots c_{k_{q-1}j},$$

which either vanishes because certain of the $c_{\alpha\beta}$'s are zero, or else represents a (possibly redundant) path of length q from i to j. Moreover, every path of length q from i to j is represented in this sum.

Now let π_{ij} denote a product $c_{ik_1}c_{k_1k_2} \cdots c_{k_{r-1}j}$ representing a proper (but not necessarily unique) path from v_i to v_j in the net. Let π_{ikj} denote a proper path from v_i to v_j passing through v_k, etc. Then we *postulate* that these path products be combined according to the laws shown in Fig. 4.

Then, since all redundancies are combinations of these three kinds, if a product representing a proper path is added to a product representing a redundant path containing that proper path, the sum reduces to the symbol for the proper path.

Evidently C is the matrix of proper paths of length 1. C^2 is the matrix of *all* paths of length 2. The sum C^2+C must then be the matrix of all proper paths of length ≤ 2 by the results of the preceding paragraph. Proceeding thus, we obtain finally the following theorem.

Theorem 4: Exactly the proper paths of length $\leq q$ of a net with connection matrix C are given by the matrix

$$C_q = \sum_{k=1}^{q} C^k.$$

The main diagonals of C_q list the proper cycles of length $\leq q$. Since a *directed Hamilton line* is defined as a proper cycle of length n, *the directed Hamilton lines of the net will be the diagonal terms with n factors in C_n.*

Now let P_q denote the matrix whose entries p_{ij} represent, as sums of products of q factors, all the proper paths, of length exactly q, from i to j in a net with connection matrix C. Since any proper path of length q, $q \geq 2$, may be regarded as a combination of a branch and a proper path of length $q-1$, we have an alternative and more practical rule which avoids many complex redundancies by eliminating in advance the simpler redundancies from which they arise.

Theorem 5: $P_1 = C$; P_q may be found from the product

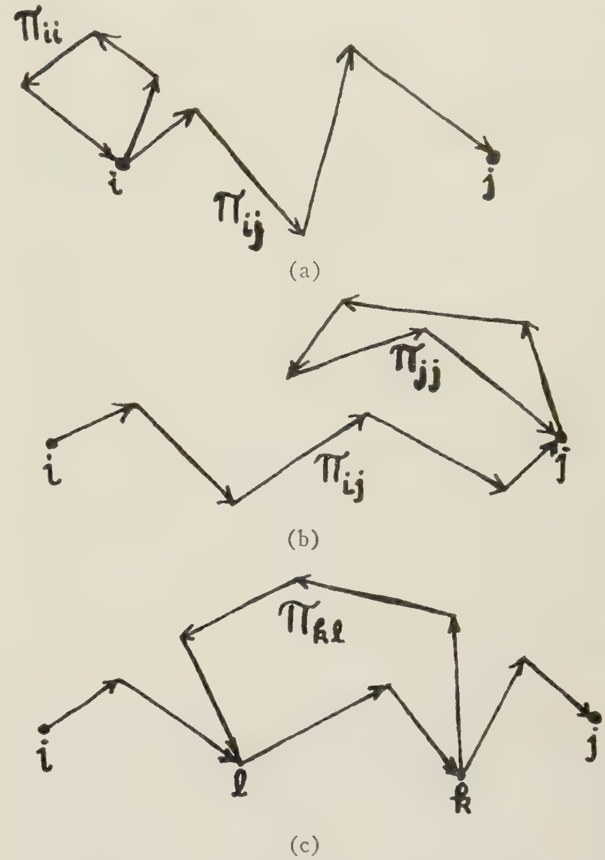


Fig. 4—(a) $\pi_{ij}+\pi_{ii}\pi_{ij}=\pi_{ij}$, (b) $\pi_{ij}+\pi_{ij}\pi_{jj}=\pi_{ij}$, (c) $\pi_{ilkj}+\pi_{ik}\pi_{kl}\pi_{lj}=\pi_{ilkj}$.

$P_{q-1}C$ by rejecting as they appear all terms in which any subscript appears more than twice.

Finally, since there are only m distinct subscripts available, any nonvanishing product $c_{k_1k_2}c_{k_2k_3} \cdots c_{k_{q-1}k_q}$ with $q > m$ must necessarily involve some subscript more than twice and hence must represent a redundant path. (Another way to see this is to note that any proper path from i to j can pass through at most $m-2$ intermediate vertices if $i \neq j$ and at most $m-1$ intermediate vertices if $i=j$.) If p is the length of the longest proper path in C, we have then $p \leq m$. We may therefore state the following result.

Theorem 6: For a given net, there exists a unique positive integer $p \leq m$ (p is the length of the longest proper path in the net) such that

$$C_{p+q} = C_p = \sum_{k=1}^{p} C^k = \sum_{k=1}^{p} P_k$$

for all positive integers q.

As an example of how these theorems are applied, consider the graph in Fig. 5.

We have

$$C = \begin{bmatrix} c_{11} & c_{12} & 0 \\ c_{21} & 0 & c_{23} \\ c_{31} & c_{32} & 0 \end{bmatrix}.$$

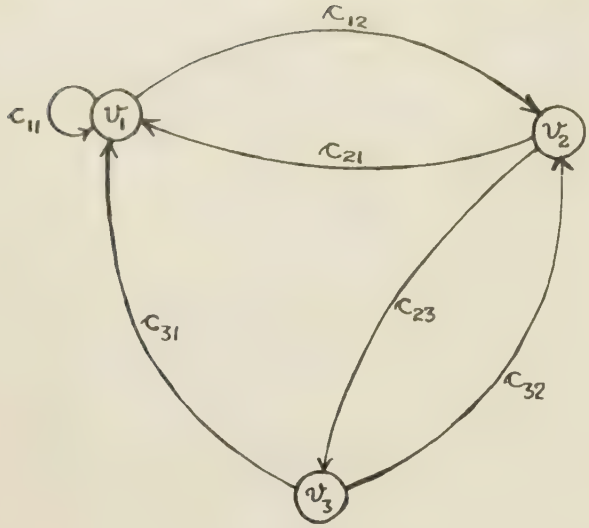


Fig. 5.

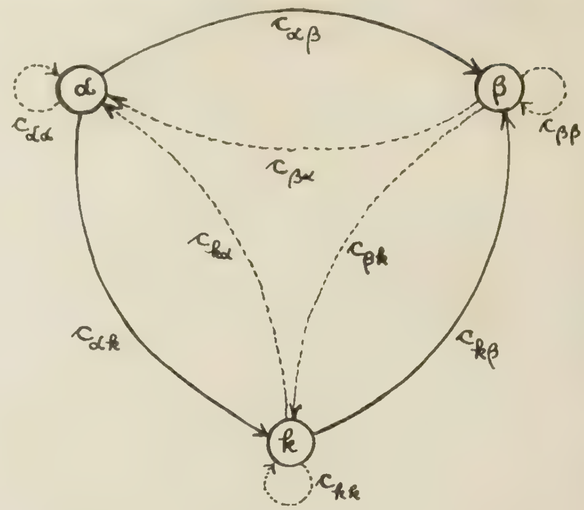


Fig. 6.

from which, readily,

$$P_2 = \begin{bmatrix} c_{12}c_{21} & 0 & c_{12}c_{23} \\ c_{23}c_{31} & (c_{21}c_{12} + c_{23}c_{32}) & 0 \\ c_{32}c_{21} & c_{31}c_{12} & c_{32}c_{23} \end{bmatrix}$$

and

$$P_3 = \begin{bmatrix} c_{12}\cdot c_{23}\cdot c_{31} & 0 & 0 \\ 0 & c_{23}\cdot c_{31}\cdot c_{12} & 0 \\ 0 & 0 & c_{31}\cdot c_{12}\cdot c_{23} \end{bmatrix}.$$

Note that P_m is always the matrix of the Hamilton lines of the net, a fact which is illustrated here by P_3.

There is a simple algorithm for computing the proper paths joining two given vertices v_i and v_j. We note first that all such paths are given by the simplified form of the sum

$$\gamma_{ij} = c_{ij} + \sum_k c_{ik}c_{kj} + \sum_{k_1,k_2} c_{ik_1}c_{k_1k_2}c_{k_2j} + \cdots$$

terminating with the sum of products of $m-1$ factors if $i \neq j$ and with the sums of products of m factors if $i=j$. The summations need to be extended only over sets of subscripts which are all distinct since only proper paths are desired.

This sum may be constructed from the connection matrix C by the following procedure, which involves the successive "removal" of vertices other than i and j. If we wish to begin by removing vertex k, we first add the product $c_{\alpha k}\cdot c_{k\beta}$ to each entry $c_{\alpha\beta}$ of C, thereafter deleting row k and column k from the matrix.

Possible connections such as $c_{k\alpha}$, $c_{\alpha\alpha}$, etc., shown dotted in Fig. 6, have nothing to contribute to proper paths from α to β. Thus any proper path or part of a path involving vertex k is still accounted for.

Starting with the resulting matrix, we remove another vertex by the same procedure and continue thus until we obtain a second-order matrix of the form

$$\begin{bmatrix} \mu_{ii} & \gamma_{ij} \\ \gamma_{ji} & \mu_{jj} \end{bmatrix}$$

which provides the desired sum γ_{ij}.

In this process we simplify as we go along, since this often forestalls the appearance of many symbols for more complex types of redundant paths which might otherwise be encountered.

If we now remove vertex v_j in the same manner, the end result is simply a single expression γ_{ii}. It includes the term c_{ii} if there is a loop at v_i. This is acceptable since such a loop is by definition a proper path. However, no loops at other vertices than v_i appear in any of these terms.

To illustrate the process, consider the net shown in Fig. 7. Denoting the result of removing node 3 by $C_{(3}$, we have

$$C_{(3} = \begin{bmatrix} (c_{11} + c_{13}c_{31}) & c_{12} \\ (c_{21} + c_{23}c_{31}) & c_{22} \end{bmatrix}$$

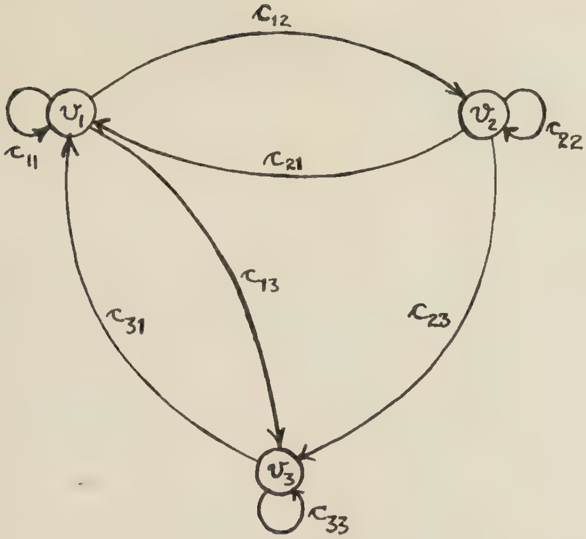
from which $\gamma_{12} = c_{12}$ and $\gamma_{21} = c_{21} + c_{23}c_{31}$. Now removing node 2, we have

$$C_{(3(2} = \gamma_{11} = c_{11} + c_{13}c_{31} + c_{12}c_{21} + c_{12}c_{23}c_{31},$$

which lists the proper cycles from v_1 back to v_1 again.

The fact that this algorithm does not include loops at vertices other than v_j is a disadvantage in certain applications.

The results of this section may be applied to undirected graphs by the simple expedient of putting $c_{ij} = c_{ji}$ in every case. The matrix C is then symmetric, which simplifies the work to some extent, but the basic facts and procedures remain the same.



$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & 0 & c_{33} \end{bmatrix}$$

Fig. 7.

Conclusion

The generality and the wide applicability of the idea of a net are an indication of its importance as a unifying concept. In succeeding papers we shall use the results given here to develop our matrix calculus for the theory of sequential machines.

Acknowledgment

The authors wish to thank the Bell Telephone Laboratories and the College of Engineering, University of Illinois, for support during various stages of this research.

Bibliography

[1] Veblen, O. *Analysis Situs.* New York: American Mathematical Society, 1931.
[2] König, D. *Theorie der endlichen und unendlichen Graphen.* New York: Chelsea Publishing Company, 1950.
[3] Copilowish, I. M. "Matrix Development of the Calculus of Relations." *Journal of Symbolic Logic*, Vol. 13 (December, 1948), pp. 193–203.
[4] Birkhoff, G. *Lattice Theory.* New York: American Mathematical Society, 1948.
[5] Hohn, F. E. "Some Mathematical Aspects of Switching." *American Mathematical Monthly*, Vol. LXII (February, 1955), pp. 75–90.
[6] Hohn, F. E., and Schissler, L. R. "Boolean Matrices and the Design of Combinational Relay Circuits." *Bell System Technical Journal*, Vol. 34 (January, 1955), pp. 177–202.
[7] Hohn, F. E. "A Matrix Method for the Design of Relay Circuits." IRE Transactions, Vol. CT-2 (June, 1955), pp. 154–161.
[8] Moore, E. F. "Gedanken—Experiments on Sequential Machines" in *Automata Studies.* Princeton, N. J.: Princeton University Press, 1956, pp. 129–153.
[9] Shimbel, A. "Applications of Matrix Algebra to Communication Nets." *Bulletin of Mathematical Biophysics*, Vol. 13, No. 3 (1951), pp. 165–173.
[10] Shimbel, A. "Structural Parameters of Communication Networks." *Bulletin of Mathematical Biophysics*, Vol. 15, No. 4 (1953), pp. 501–507.
[11] Shimbel, A. "Structure in Communication Nets." *Proceedings of the Symposium on Information Networks*, Polytechnic Institute of Brooklyn (1954), pp. 199–203.
[12] Mason, S. "Feedback Theory—Some Properties of Signal Flow Graphs." Proceedings of the IRE, Vol. 41 (February, 1953), pp. 1144–56.
[13] Mason, S. "Feedback Theory—Further Properties of Signal Flow Graphs." Proceedings of the IRE, Vol. 44 (December, 1956), pp. 920–926.
[14] Luce, R. D., and Perry, A. D. "A Method of Matrix Analysis of Group Structure." *Psychometrika*, Vol. 14 (March, 1949), pp. 95–116.
[15] Luce, R. D. "Connectivity and Generalized Cliques in Sociometric Group Structure." *Psychometrika*, Vol. 15 (June, 1950), pp. 169–190.
[16] Polya, G. "Sur le nombre des isomères de certains composés chimiques." *Comptes Rendus*, Vol. 202 (May, 1936), pp. 1554–1556.
[17] Hertz, P. "Über Axiomensysteme für Beliebige Satzsysteme." *Mathematische Annalen*, Vol. 87 (1922), pp. 246–269.
[18] Percival, W. S. "Solution of Passive Electrical Networks by Means of Mathematical Trees." *Proceedings of the Institution of Electrical Engineers*, Vol. 100, Part III (May, 1953), pp. 143–150.
[19] Reed, M. B., and Seshu, S. "On Topology and Network Theory." *Proceedings of the University of Illinois Symposium on Circuit Analysis* (1955), pp. 2–1 to 2–16.
[20] Luce, R. D. "Two Decomposition Theorems for a Class of Finite Oriented Graphs." *American Journal of Mathematics*, Vol. LXXIV (July, 1952), pp. 701–722.
[21] Luce, R. D. "Networks Satisfying Minimality Conditions." *American Journal of Mathematics*, Vol. LXXV (October, 1953), pp. 825–838.
[22] Mealy, G. H. "A Method for Synthesizing Sequential Circuits." *Bell System Technical Journal*, Vol. 34 (September, 1955), pp. 1045–1079.
[23] Burks, A. W., and Wang, H. *The Logic of Automata.* Ann. Arbor, Mich.: University of Michigan, Engineering Research Institute, December, 1956.
[24] Burks, A. W., and Wright, Jesse B. "Theory of Logical Nets." Proceedings of the IRE, Vol. 41 (February, 1953), pp. 1357–1365.
[25] Huffman, D. A. "The Synthesis of Sequential Switching Circuits." *Journal of the Franklin Institute*, Vol. 257 (March and April, 1954), pp. 161–190, 275–303.
[26] Turing, A. M. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings London Mathematical Society*, Vol. 42 (November, 1937), pp. 230–265. Corrections in Vol. 43 (May, 1937), pp. 544–546, and in reference [27].
[27] Post, Emil L. "Recursive Unsolvability of a Problem of Thue." *Journal of Symbolic Logic*, Vol. 12 (March, 1947), pp. 1–11.
[28] Reed, I. S. *Some Mathematical Remarks on the Boolean Machine.* Lexington, Mass.: Massachusetts Institute of Technology, Lincoln Laboratory, Technical Report No. 2, August, 1952.

A Design Technique for Pedestal-Free Switching Circuits*

GEORGE SEBESTYEN†

Summary—The fundamental differences between mechanical and electronic switching circuits are discussed with special emphasis on the problem of pedestals. Based on the fundamental differences it is shown that Boolean algebra fails to express practical requirements of electronic switches.

This paper takes a step in the direction of removing this limitation by augmenting the ON-OFF notation of Boolean algebra by a notation of inequalities to handle the problem of pedestal-free switching.

The ideal electronic relay is defined and a diode network realization is synthesized. An electronic switch is synthesized to illustrate the design technique's application. Photographs showing the performance of the circuits are given.

WITH the advent of general purpose computers and other large scale devices, it has become increasingly important to take the art out of the design of switching circuits and invoke the use of a systematic approach made possible through the medium of Boolean algebra. The end result of this latter method of attack is a network topology of interconnected contacts, some normally open, others normally closed.

At this point, the theoretical work comes to an end and the engineer is faced with the difficulty of translating his network of contacts into a fast-acting electronic switching circuit. The principal obstacle in the path of the transition from the network of contacts to its electronic counterpart is the failure of electronics to supply the ideal building block instrumented so perfectly by a relay. Stating this differently, the obstacle is the failure of Boolean algebra in its ideal form to express practical requirements of electronic circuits with the precision with which it can express the actual operation of mechanical switches. The object of this paper is to take a step in the direction of removing this limitation by augmenting the ON-OFF notation of Boolean algebra by a notation of inequalities to handle the problem of pedestal-free switching in electronic circuits.

In strictly digital switching, the lack of an ideal building block does not prove to be a serious defect, for the digital switch is required to transmit no more information than that the switch is ON or OFF. In the switching of analog quantities or continuous functions, however, the switch requirements are much more severe. Instead of merely stating that it is now ON or OFF, the switch must transmit or discontinue to transmit the flow of arbitrary information. The requirement on the part of the electronic switch to provide only a continuity or a discontinuity of the flow of information is hindered by the unwanted appearance of the switching signal commanding the circuit to change from one state to another. This unwanted switching signal in the output is sometimes called the pedestal, because it is a dc signal or a pulse on which the information appears superimposed as if on a pedestal. The presence of this pedestal makes most electronic switches fall short of the goal for finding an exact high-speed equivalent of the relay.

In many applications the switching signal can be eliminated by conventional filtering. This is possible where the principal portion of the signal spectrum and the spectrum of the switching signal do not overlap. Most often, however, simple filtering will not provide a suitable solution—at least it is not a general solution that gets to the root of the problem.

Properties of Mechanical and Electronic Switches

The principal building block of switching circuits is the ideal relay shown in Fig. 1(a). It is used as a two-terminal device which is a short circuit between terminals 1 and 2 if the energizing coil a is suitably excited, and it is an open circuit between terminals 1 and 2 when a is not energized. These two states are the 1 and the 0 states of the contact A, respectively. Although the device itself has only two terminals, it is almost always associated with a four-terminal application in analog switching. In the four-terminal network [Fig. 1(b)], the relay will transmit the information E_1 if $A=1$ and will not transmit it to the output if $A=0$. The two-terminal pair device [Fig. 1(b)] is a better, practical representation of a relay contact as it is used. After all, relays are used to transmit or interrupt the flow of information and information in electronics always appears across a pair of terminals. Since it is convenient to take the input and output voltages with respect to the same potential, and since this does not result in a loss of generality, it is proper to represent a practical relay contact as the transfer function implied by the circuit of Fig. 1(b). The transfer characteristics of the ideal relay of Fig. 1(b) are listed below:

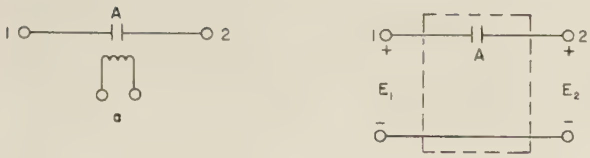
$$E_2 = E_1 \quad \text{if } A = 1;$$

$$E_2 = \text{undetermined or floating if } A = 0.$$

This latter characteristic of a switch, that the two terminals of the output have no definite relationship to each other if the switch is open, is a unique property of switches brought about by the mechanical separation of the input and output terminals which takes place when a relay contact is broken. No such break occurs in an

* Manuscript received by the PGEC, January 25, 1957; revised manuscript received, May 10, 1957.

† Research Dept., Melpar, Inc., Boston, Mass.



(a) (b)

Fig. 1—Ideal relays.

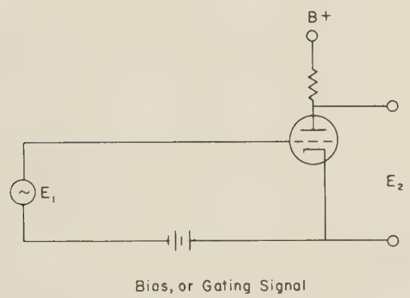
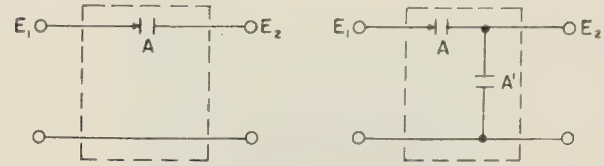
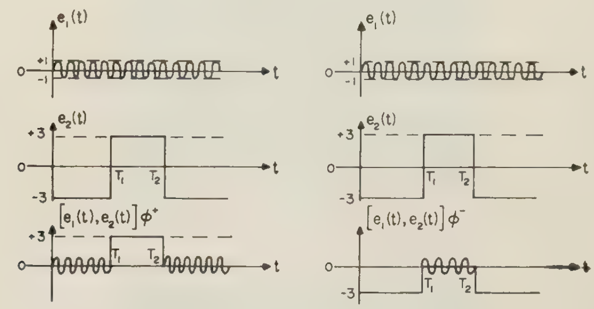


Fig. 2—Vacuum tube switch.



(a) (b)

Fig. 3—Ideal switch and its electronic counterpart. (a) $A=0$, $E_2=?$, $A=1$, $E_2=E_1$; (b) $A=0$, $E_2=0$; $A=1$, $E_2=E_1$.



(a) (b)

Fig. 4—Illustration of ϕ^+ and ϕ^- operations.

electronic network with no moving parts, and consequently, terminals across which information may appear at one time must be fixed and cannot float with respect to each other at other times.

Consider the vacuum tube circuit of Fig. 2, for example. When the grid is biased in the linear amplification region of the tube, information can pass from the grid to the plate circuit, and the tube may be regarded incrementally as an energized relay. When the grid is biased highly negative, the tube is at cutoff and transmits no information to the plate circuit. In this respect, a cutoff vacuum tube may be thought of as an open switch. In reality, however, it is far from being that. As pointed out above, the output voltage of an open switch is undetermined. The output of a cutoff vacuum tube amplifier is fixed at some potential such as the potential of the plate supply voltage in the example above. The situation is similar if devices other than vacuum tubes are considered as the switched elements. For this reason, it is practical to redefine the meaning of a switch as used in electronics by returning the output to a fixed reference, such as ground, in the unenergized state. It may be stated, therefore, that the output of an ideal electronic switch is equal to the input when the switch is ON and is identically equal to zero when the switch is OFF. Fig. 3(a) is the ideal switch whose electronic counterpart is now defined as a circuit realizing the grounded transfer contact network shown in Fig. 3(b). This definition is a particularly reasonable one in view of the constant strife for building better switches which exhibit no pedestals.

Some Operations in an Algebra of Inequalities

Before the synthesis of circuitry can be undertaken, a departure from the main subject must be made to discuss some basic algebraic operations to be used later on. The two operators to be used are symbolic representations of the words "most positive" or "greatest" and the words "most negative" or "least." It is beyond the scope of this article to discuss all the interesting properties of an entire algebra of inequalities built around these two operations.[1] Here, use is made only of the two above mentioned operations denoted by $(\ )\phi^+$ and $(\ )\phi^-$. The operator $(\ )\phi^+$ selects the greatest or most positive member of a set of quantities appearing as its argument. The operator $(\ )\phi^-$, on the other hand, selects the least or most negative member of the set of quantities comprising its argument. To illustrate the usefulness of these operators in the analytical description of switching circuits, consider the following example. Of the two time functions shown in Fig. 4(a), $e_1(t)$ is a sine wave of unity amplitude and $e_2(t)$ is a square wave of amplitude three. If $e_1(t)$ and $e_2(t)$ are made the argument of the $(\ )\phi^+$ operator, written as $[e_1(t), e_2(t)]\phi^+$, then the output is seen to be the sine wave $e_1(t)$ from $t=0$ to $t=T_1$ and it is the square wave $e_2(t)$ from $t=T_1$ to $t=T_2$. Similarly, if the operation $[e_1(t), e_2(t)]\phi^-$ were performed to select at all times the most negative of $e_1(t)$ or $e_2(t)$, then the output would be the square wave for $t<T_1$, and the sine wave for $T_1<t<T_2$, as shown in Fig. 4(b). The application to switching circuits is now quite evident. The square wave may be regarded as the switching or gating signal and the sine wave as the information to be switched. The only requirement for successful switching is that the peak amplitude of the information signal be always less than the gating signal. This is easily accomplished in practice. The circuits which can accomplish the two basic operations are shown in Fig. 5. The circuit of Fig. 5(a) is the instru-

[1] T. B. Stern, "Piecewise-Linear Network Theory," Sc.D. Dissertation, Mass. Inst. Tech., Cambridge, Mass.; June, 1956.

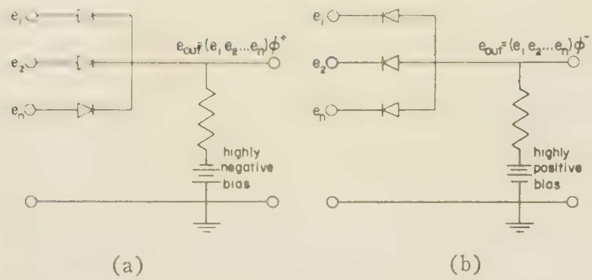


Fig. 5—The ϕ^+ and ϕ^- operators.

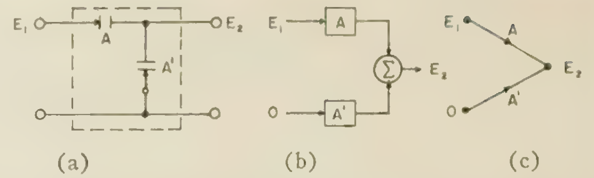


Fig. 6—Unidirectional ideal electronic switch with block diagrammatic and signal flow graph representation. (a) $E_2=0$ if $A=0$, $E_2=E_1$ if $A=1$; (b) $E_2=E_1A+OA'$; (c) $E_2=E_1A+OA'$.

mentation of the ϕ^+ operator. The diode whose anode is most positive will conduct and will be a short circuit, thus causing the output to equal its anode voltage. Similarly, the circuit of Fig. 5(b) realizes the ϕ^- operator because the diode whose cathode is most negative will conduct, causing the output to assume the conducting diode's cathode potential. All other diodes are nonconductive, thus providing isolation of inputs. These circuits are suggested to instrument the two basic operators.[1] The ϕ^+ and the ϕ^- circuits are seen to be identically the same as the logical *or* and the logical *and* circuits of conventional digital switching circuits. This is quite reasonable if the ϕ^+ operation is given a strictly two-valued interpretation for use in Boolean algebra. The operation of addition in Boolean algebra means, for instance, that if any one or more of the digits A, B, $C, \cdots$ appearing in the sum $(A+B+C+\cdots)$ is one, the sum is one. This is the logical *or*. Of the two values, 0 and 1, which the digits may assume, Boolean addition selects the highest value present as the sum. This is exactly what the ϕ^+ operator does. In Boolean multiplication, however, if any one or more of the digits is zero, the product $(A\times B\times C\cdots)$ is zero. This is the logical *and* requiring A and B and C and all the factors of the product to be 1 for the entire product to have the value unity. Of the two values, 0 and 1, which the digits may assume, Boolean multiplication selects the lowest value present as the value of the product. This is exactly the function performed by the ϕ^- operator. It may be stated, therefore, that:

$$(A+B+C+\cdots) = (A, B, C, \cdots)\phi^+ = \text{logical } or \text{ and that}$$

$$(A\times B\times C\times\cdots) = (A, B, C, \cdots)\phi^- = \text{logical } and.$$

These two equations hold true not only if the argument is 0 or 1 but also if it has highly positive or highly negative values. Highly positive or negative voltages are more meaningful when talking about analog switching, where instead of having closed relay contacts, we might have highly positive gating signals, and instead of open relays, we might have highly negative cutoff voltages. Although stretching the point a bit, the equivalent meaning of the Boolean operations and the ϕ^+ and ϕ^- operators, as will be seen, is useful not only if the individual members in the set A, B, $C, \cdots$ have two discrete values, but also if they are contained in nonoverlapping regions.

Having introduced the above two equations, the synthesis of the ideal electronic switch of Fig. 3(b) may be resumed. In the case of practical electronic switching circuits, the switches used are intended to be unidirectional. The switch is to have not only two pairs of terminals, but a set of noninterchangeable input and output terminal pairs. This implies that information may flow from input to output but not from output to input. In this respect, the electronic equivalent of a relay contact functions the same way as blocks in a block diagram or transmissions in a signal flow graph, both of which operate on their input quantities. For this reason, the ideal unidirectional electronic switch of Fig. 6(a), where the arrows indicate the permissible direction of the flow of information, can be represented by the block diagram or flow graph of Fig. 6(b) and 6(c), respectively. In the equations describing the block diagram and flow graph, the second input of 0 is not superfluous. It represents the magnitude of the required fixed voltage output of the switch in the OFF state. This does not necessarily have to be zero. It was defined to be zero merely because no pedestal was desired in the output.

As a result of the preceding discussion, we were able to express the output of a switch as the sum of two products. Applying the previously derived equivalence of ϕ^+ operation to addition and ϕ^- operation to multiplication, the following equation is obtained:

$$E_2 = [(E_1, A)\phi^-, (O, A')\phi^-]\phi^+.$$

The term $(E_1, A)\phi^-$ describes the product E_1A and the term $(O, A')\phi^-$ denotes the product of OA'. The sum of these two terms is taken by the ϕ^+ operator to yield the above equation which can be instrumented by the suitable interconnection of ϕ^+ and ϕ^- operators as shown in Fig. 7.

In Fig. 7, E_1 is the signal to be switched and A and A' are the switching signal and its complement, respectively. If $A=0$, then the switching signal is highly negative and its complement highly positive; if $A=1$, the switching signal is highly positive and its complement highly negative, as stated earlier. When the electronic switch is to be OFF, the $\phi^-_{(1)}$ network selects the more negative of the information and the highly negative switching signal. At the same time the $\phi^-_{(2)}$ network selects the zero input because that is lower than the switching signal's complement. Of the outputs of the

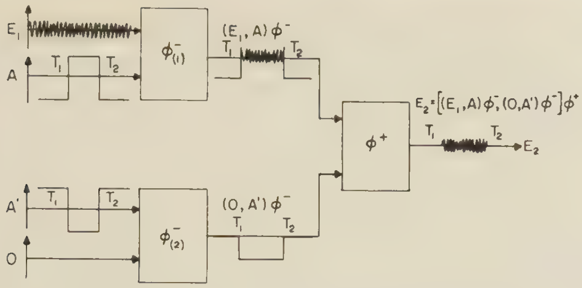


Fig. 7—Electronic realization of the switch.

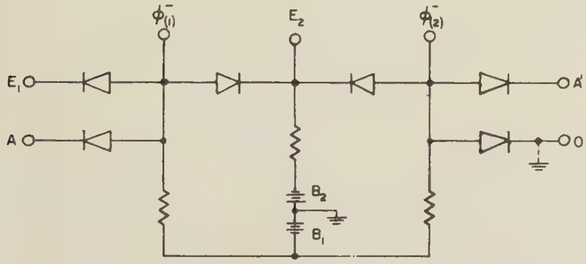
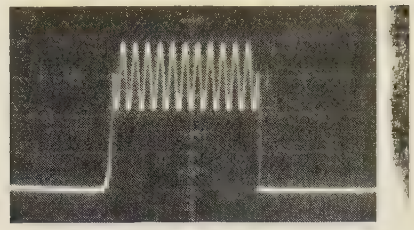
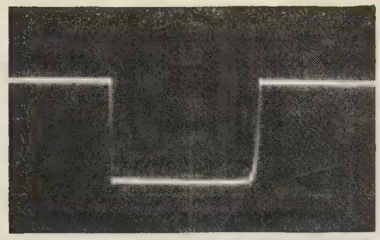


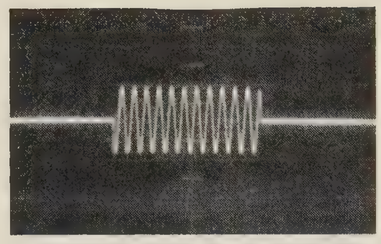
Fig. 8—Network realization of ideal electronic switch.



(a)



(b)



(c)

Fig. 9—Waveforms in the ideal switch. (a) $\phi^-{}_{(1)}(t)$; (b) $\phi^-{}_{(2)}(t)$; (c) $[\phi^-{}_{(1)}, \phi^-{}_{(2)}]\phi^+$.

two ϕ^- networks, the zero output of $\phi^-{}_{(2)}$ is greater. Hence, the ϕ^+ operator selects zero as its output. As the gating signal is turned ON, the $\phi^-{}_{(1)}$ operator passes the information signal E_1 because that is the more negative input. The $\phi^-{}_{(2)}$ operator, on the other hand, passes the highly negative complement of the gating signal. During the ON period, therefore, E_1 is the more positive of the two ϕ^- operator outputs and is passed on to the final output by the ϕ^+ operator.

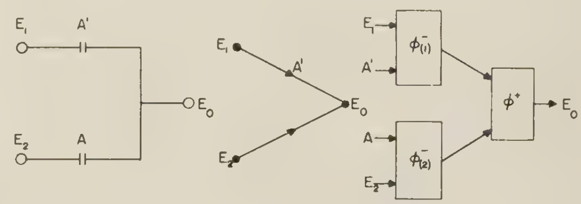
With this interconnection of ϕ^- and ϕ^+ operators, governed by the equation derived from the definition of an ideal unidirectional electronic switch, all the initially imposed requirements have been met. The signal appears with unity gain and without the objectionable presence of a pedestal.

The network which realizes the block diagram of Fig. 7 is shown in Fig. 8. The bias voltage B_1 is common to both ϕ^- circuits. All input terminals must have a dc return to ground and should be driven from low impedance sources. Waveforms at various points in the circuit are shown in Fig. 9, and are found to agree with the theory as illustrated in Fig. 7. A sine wave is used as the information signal.

Circuit Application

Although a unified theory for the synthesis of analog switching circuits with the above techniques has not yet been completed, the design of a simple switching circuit with the above methods would bring out the utility of a systematic method of attack.

Consider the realization of an electronic switch used to display the sequential alternation of two input functions on a single cathode-ray tube. The required contact network is shown in Fig. 10(a), where A switches alternatingly between 0 and 1 and A' between 1 and 0. The first step is to draw a signal flow graph, shown in Fig. 10(b), from which the network equation may be written. The network equation may now be transformed into the $\phi^\pm$ notation with the aid of previously developed equations. This yields

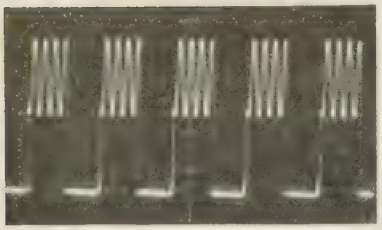


(a) (b) (c)

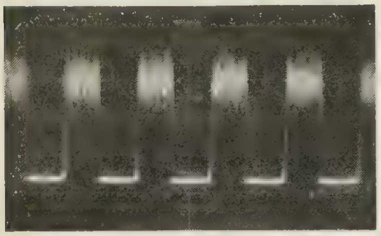
Fig. 10—Electronic switch. (a); (b) $E_0 = E_1A + E_2A$; (c) $E_0 = [(E_1, A')\phi^-, (E_2, A)\phi^-]\phi^+$.

$$E_0 = E_1A' + E_2A = [(E_1, A')\phi^-, (E_2A)\phi^-]\phi^+$$

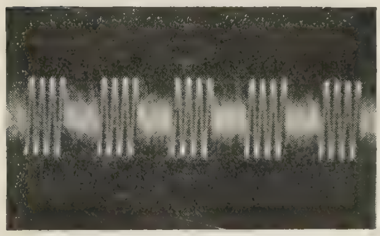
realized by the block diagram of Fig. 10(c). This diagram is recognized to be identical to that of Fig. 7 with the exception that instead of applying 0 to the $\phi^-{}_{(2)}$ input, the second function E_2 is applied. The photographs of Fig. 11 show the waveforms when E_1 is a sine wave and E_2 is random noise. The switching functions are obtained from opposite plates of a multivibrator through a coupling network to obtain a positive as well as negative going square wave-switching signal.



(a)



(b)



(c)

Fig. 11—Waveforms. (a) $\phi^-_{(1)}$; (b) $\phi^-_{(2)}$; (c) $[\phi^-_{(1)}, \phi^-_{(2)}]\phi^+$.

amplitude of the information to be switched.

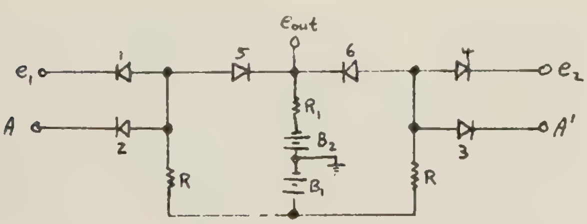
It should be noted that switches which operate on this principle also may be instrumented using analog computer operational amplifier.[2]

In addition, by an analysis of the switch, it is shown in the Appendix that the crosstalk—the ratio of the signal magnitude at the output in the OFF state to that in the ON state—can be made close to $2(R_f/R_r)^2$, where R_f is the forward and R_r is the reverse resistance of the diodes used in the circuit. In practice, this may be a very impressively and unmeasurably small crosstalk.

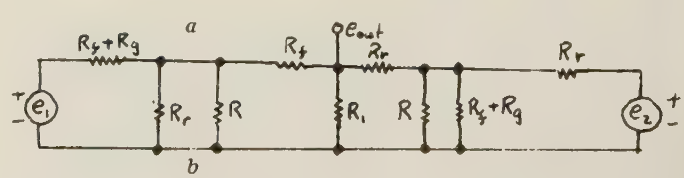
Appendix

Proof That Crosstalk Can Be Made Close to $2(R_f/R_r)^2$

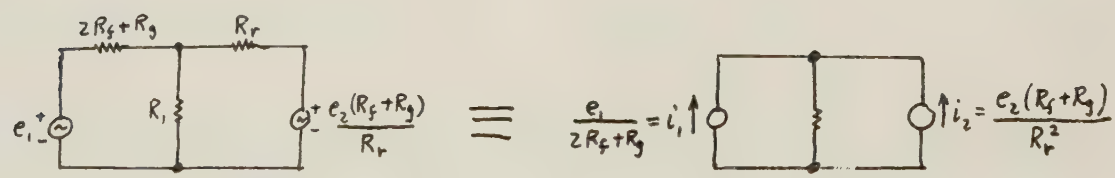
Crosstalk is the ratio of the signal gain in the OFF state to that in the ON state. In the circuit below, when A is highly positive and A' highly negative, crosstalk is the ratio of the gain (or attenuation) from e_2 to e_{out} to the gain from e_1 to e_{out}. Diodes 1, 3, and 5 are conducting and 2, 4, and 6 are not [Fig. 12(a)]. Diodes conducting have resistance R_f, diodes in the OFF state have resistances R_r. Each voltage generator has an internal impedance R_g. The ac incremental circuit is shown in Fig. 12(b). Assumptions made are that R, $R_2 \gg R_g$, R_f. If the parallel combination of R_r and R is very large compared to R_f+R_g, then the branch consisting of R_r in parallel with R between points ab may be neglected.



(a)



(b)



(c)

Fig. 12.

Conclusion

In the foregoing, synthesis of a pedestal-free electronic switch was described prompted by some algebraic inequality operations. The switch in its final form is a simple device and is capable of very-high-speed operation. The principal advantage of the device is that by its very nature, neither the magnitude nor the shape of the switching function are of any consequence insofar as successful uniform switching is concerned. In fact, as stated earlier, the only required property the switching signal must possess is that it be greater than the peak

Similarly, the right-hand side of the circuit may be simplified to look like the circuit in Fig. 12(c). These simplications have made use of the valid assumptions stated above. The crosstalk is the ratio

$$i_2/i_1 = \frac{(R_f + R_g)(2R_f + R_g)}{R_r^2}.$$

If the generator impedance can be made very small, the crosstalk would become $2(R_f^2/R_r^2)$, as stated in the text.

[2] R. M. Howe, "Representation of nonlinear functions," IRE Trans., vol. EC-5, pp. 203–206; December, 1956.

A New Method for Generating a Function of Two Independent Variables*

LAZARUS G. POLIMEROU†

Summary—A new concept in the design of a function generator of two independent variables is presented. This design is of the photoformer variety and utilizes pulse techniques to obtain the advantages of accuracy and semiautomatic calibration and self-correction. With this particular design, many of the objections of other types of photoformers, which have been used to generate functions of single independent variables, have been eliminated, and yet this design maintains the desirable aspects of the photoformer.

Introduction

THE generation of an arbitrary function of two or more variables is an important computer operation. The nature of problems being solved on present analog computers is complicated by the need for generating such functions. Available methods for meeting such needs are well outlined by Meissinger.[1] The author now proposes an additional design which has the very important feature of simplifying the set-up procedure because the function generator is of the photoformer type and utilizes pulse techniques to obtain improved operation. This design is an extension of a basic experimental model containing four units, completed January, 1953.[2] A model based on the extended design is now under investigation.[3]

A Function of One Independent Variable

The basic function generator[4] is composed as follows: an electron beam of a crt is made to scan a mask (as shown in Fig. 1) at a relatively high sweep rate. As the beam passes the reference line of the mask, a pulse is produced by a photomultiplier tube which looks only at this reference line. The beam continues to move and scans past the function line of the mask, whereupon a second pulse from a second photomultiplier tube, looking only at the function line, is produced. The time delay between these two pulses is directly proportional to the length of the ordinate between the reference line and the function line, assuming a linear sweep. These two pulses are then shaped and fed into grids of a flip-flop circuit which produces a square wave that has a width proportional to the time delay between the pulses and that has a constant amplitude. In this process, the first pulse initiates the square wave and the second pulse terminates the square wave. Thus a train of square waves is produced. The square waves are then rectified and integrated, thus producing a varying dc voltage that is proportional to the ordinate between the reference line and the function line; *i.e.*, a voltage proportional to the value of the function at the point being scanned.

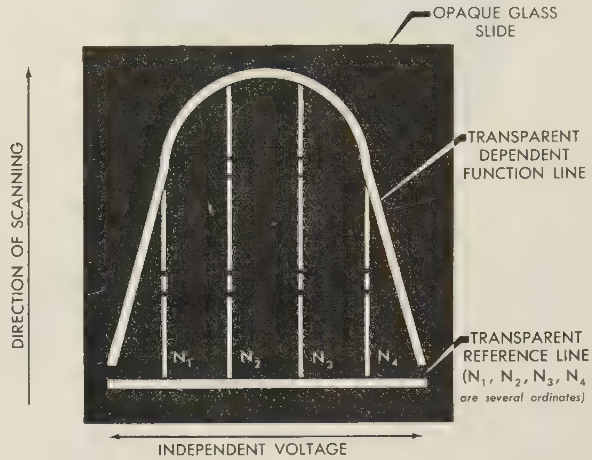


Fig. 1—Function slide for function of one independent variable.

As the ordinate length varies, so does the output voltage. By moving the scanning line (in a direction perpendicular to the scan, called the independent function axis) over the mask which represents the dependent function, the ordinate lengths can be made to vary with respect to the position of the scan. If the scanning line is moved in a linear fashion with respect to time along the independent axis, a voltage is produced in the output which is the electrical replica of the mask, as given in Fig. 1. Furthermore, the scanning line can be moved in a nonlinear manner along the independent function axis, thus producing a function of a function.

A Function of Two Independent Variables

Having examined the basic design, a further application of this design is considered. Referring to the original slide (Fig. 1), the function line is replaced by a whole family of curves (Fig. 2). The same reference line is maintained.

Now consider the scanning point of light as it crosses the slide. When it first crosses the reference line, a pulse is produced by a photomultiplier tube which, when shaped, is used to initiate a square wave in a number of parallel flip-flop circuits (Fig. 3). The scanning point continues to scan across the slide until it crosses the first

* Manuscript received by the PGEC, February 20, 1957; revised manuscript received, April 21, 1957. This paper is based on a thesis submitted for the M.S. degree to the Dept. of Elec. Eng., Southern Methodist University, Dallas, Texas, 1957. Patent pending.

† Convair, Fort Worth, Texas.

[1] H. F. Meissinger, "An electronic circuit for the generation of functions of several variables," 1955 IRE Convention Record, part IV, pp. 150–161.

[2] At White Sands Proving Ground, N. M.

[3] At Convair, Fort Worth, Texas.

[4] L. G. Polimerou, "A new method of generating functions," IRE Trans., vol. EC-3, pp. 29–34; September, 1954.

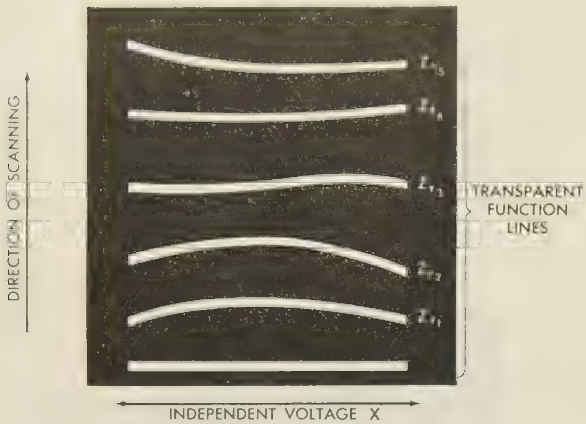


Fig. 2—Function slide for function of two independent variables.

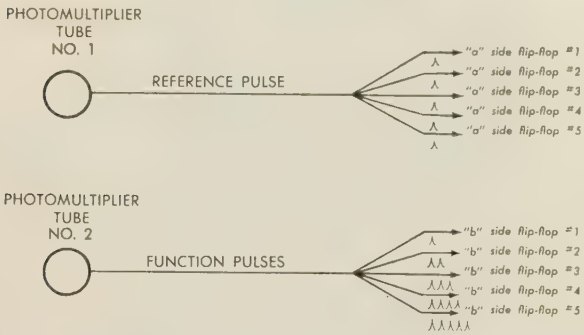


Fig. 3—Pulse distribution schematic.

function line, at which time a second photomultiplier tube produces a pulse which is sent out to all parallel flip-flop circuits. This first function line pulse terminates the square wave in flip-flop number one. However, it does not terminate any of the other square waves in the other parallel flip-flop circuits because these flip-flops are made to respond in succession according to the number of function pulses corresponding to their position on the slide. Next, as the spot of light crosses function line number two, a second function line pulse is produced which is sent out to all parallel flip-flop circuits such that this pulse terminates the square wave in flip-flop number two. Now, the second terminating pulse does not affect flip-flop number one, because the square wave at that point has already been terminated. Furthermore, it does not affect any of the other parallel flip-flop circuits, since each parallel flip-flop circuit will respond only to its own certain number of pulses corresponding to its order of succession on the slide. Similarly, it takes three pulses to terminate flip-flop circuit number three and four pulses to terminate flip-flop circuit number four, etc. Once more the retriggering of the reference pulse again initiates square waves in all parallel flip-flop circuits and the process is repeated.

Now by processing the outputs of these flip-flop circuits as in the basic design, it is possible to produce many functions on just one slide. (These functions are considered not to cross one another, although crossovers can be handled.) This result is very useful in itself. However, we have not as yet produced a function of two independent variables.

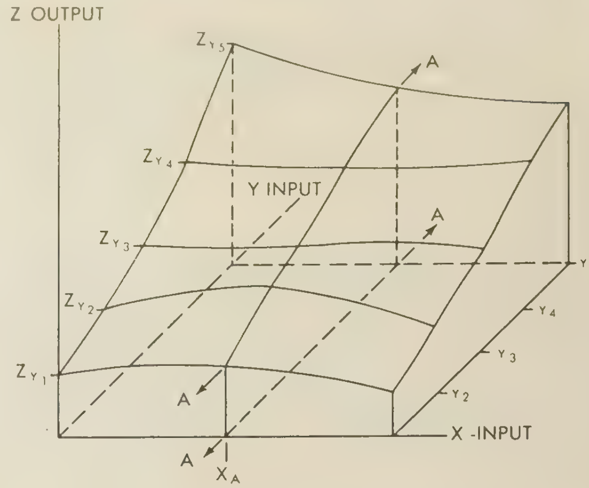


Fig. 4—Three-dimensional representation.

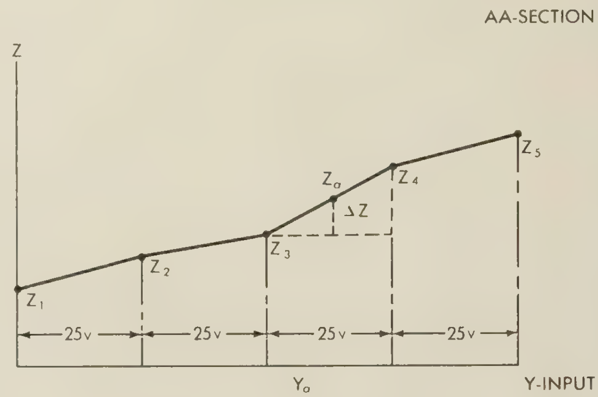


Fig. 5—Cross section of three-dimensional representation.

If switching elements are now used which will select the desired flip-flop circuit output according to a variation of a second independent variable, then the output will be a function of two independent variables. These switching elements may be high-speed relays or other electronic devices. Examining Fig. 4, it is found that the ordinate of the surface is the value Z which is selected according to both X and Y. Next, if a cross section of the surface in which X is a constant is examined (Fig. 5), the following analysis can be made. The values Z_1, Z_2, Z_3, Z_4, and Z_5 represent discrete values of Z as formed by a set of points (constant X value) on the slide (Fig. 2). Now since it is desired to have intermediate values of Z corresponding to intermediate values of Y, linear interpolation is performed. At a point $Z_a = Z_3 + \Delta Z$, corresponding to Y_a, ΔZ, is determined thusly,

$$\frac{Z_a - Z_3}{Y_a - Y_3} = \frac{\Delta Z}{Y_a - Y_3} = m,$$

m being the slope between Z_3 to Z_4 which may be expressed by

$$m_{34} = \frac{Z_4 - Z_3}{25}.$$

Therefore,

$$\Delta Z = \frac{Z_4 - Z_3}{25} \cdot (Y_a - Y_3)$$

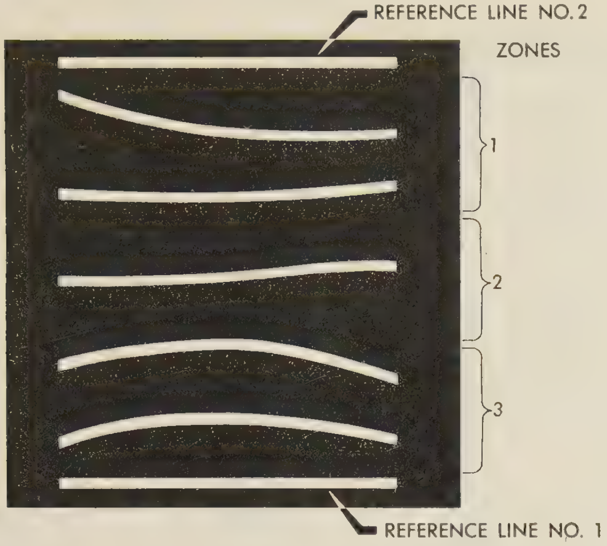


Fig. 6—Function slide for Z-axis automatic calibration.

which for circuitry is written

$$\Delta Z = \frac{1}{25}\left[Z_4(Y_a - Y_3) - Z_3(Y_a - Y_3)\right].$$

It is seen that Z_3 and Z_4 are proportional to the width of the square waves formed at the respective flip-flops concerned. Amplitude modulation of these square waves according to $(Y_a - Y_3)/25$ will produce the desired multiplication when the train of pulses from each flip-flop circuit is averaged, respectively. Consequently, the proper value of Z is selected according to both independent variables X and Y.

The following items concerning the basic design of this function generator are pointed out. First, in comparing this type of function generator with any other type photoformer, a parallel extension of the method described (in which instead of merely one function per slide being generated, many functions can be generated on a single slide) can not be duplicated on other present type photoformers.

Second, changes of beam light intensity do not affect the output of this function generator. Since if for some reason the intensity of the electron beam is increased causing an earlier triggering of the reference pulse, a similar earlier triggering will occur for the function pulse. The area under the generated square wave remains the same, therefore, the value of the function is not distorted as it would be in other types of photoformers.

Next, by the addition of a second reference line (Fig. 6), automatic calibration in the direction of scanning circuit can be had. For since the output of a second reference line (parallel to the first) should be constant, it can be set against a laboratory standard and maintained whereupon proportionate correcting voltages can be sent out to functions within any particular zone of the slide. This tends to decrease any inaccuracies due to any nonlinearities in the sawtooth scanning waves. Whereas in other types of photoformers it is necessary to set gains on both X and Z axes (thereby making it

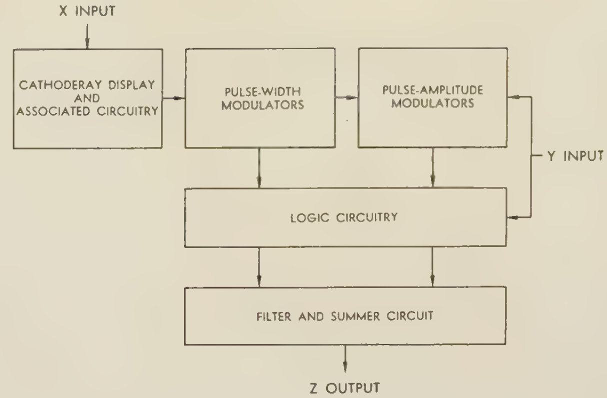


Fig. 7—Simplified system block diagram.

difficult to determine whether an error which exists in some output voltage is due to improper input gain setting or improper output gain setting or to what combination in both), such is not now the case in this function generator. In this pulse-type function generator, guessing the error in the Z direction or scanning direction (corresponding to output voltage gain) is replaced by a self-correcting and self-calibrating device.

Next, generating several functions of time on a single slide simultaneously, decreases the possibility of poor synchronization which may occur when functions are placed on a separate slide. Also, movement of the slide in the direction of scanning does not affect the output of the function generator.

Furthermore, this basic function generator design is compatible with the box-type multiplier, so that, in every case in which a unit is not used to generate a specific function, each unit can then be made available to give multiple-multiplication outputs.

In comparing this function generator to diode function generators, it has the advantages of a simple set-up and also storage of the information on the slide.

Conclusion

Features of this type system are enumerated below.

1) High accuracy—better than $\frac{1}{2}$ per cent.
2) High-frequency response—flat to beyond 100 cycles per second.
3) Elimination of curve riding errors.
4) Automatic output calibration.
5) Self-correcting scanning system.
6) Availability of multiple-multiplication channels when not generating a function of two independent variables.
7) Generation of many functions on a single slide simultaneously.
8) Simple set-up for functions to be generated.

A simplified block diagram of the system is given in Fig. 7.

Acknowledgment

The author wishes to express his appreciation for the many helpful suggestions made by H. Horn at White Sands Proving Grounds, N. M.

An Analog Method for the Solution of Probability of Hit and Related Statistical Problems*

THOMAS B. VAN HORNE†

Summary—Approximate solutions of probability of hit problems have been achieved by means of a simulation involving a masked oscilloscope and photocell for the target and samples from random noise generators for the system dispersions. The method is described in general and its possible application to other noise and statistical problems indicated. The statistical limitations on accuracy and computation rate, inherent in such a sampling method, are analyzed for an idealized probability of hit problem which may be used as a guide in choosing appropriate noise spectra for the mechanization of similar problems.

Examples which illustrate two types of restrictions on the noise spectra arising in amplifier bandwidth and screen response are considered in detail and the statistical relationship between computation rate and accuracy determined for spectra shaped by some simple RC filters.

EVALUATION of weapon system effectiveness commonly involves a calculation of the armament probability of hit by integrating the error distribution over some sort of irregular vulnerable area. In any but the simplest cases the calculation may prove tedious. One approach, which has proven successful for airborne rocket fire control systems and should be useful in many related problems, is outlined in this paper.

A sampling method[1] is employed in conjunction with electrical analogs of the system and target. The basis of the simulation is an oscilloscope with its screen covered, except for a portion in the shape of the target vulnerable area. Gaussian or other noise and constant voltages are applied to the horizontal and vertical deflection plates to represent system and rocket dispersions and biases.

With proper scaling the probability of hit is equal to the average fraction of time which the spot is within the vulnerable area. This may be determined by integrating the output of a photomultiplier tube placed in front of the screen. Alternatively, the oscilloscope intensity may be periodically pulsed and the photomultiplier output pulses counted and compared with the total number of pulses in the same interval.

The method is applicable to several other types of statistical problems. The following examples indicate some of the possibilities.

A convenient way to measure the cumulative probability distribution of a noise voltage is shown in Fig. 1. With an ideal diode the density function of the vertical input for a noise with density $p(v)$ is like that of Fig. 2. If the edges of the mask are scaled to $\pm a$ volts the probability of a sample appearing in the cutout is

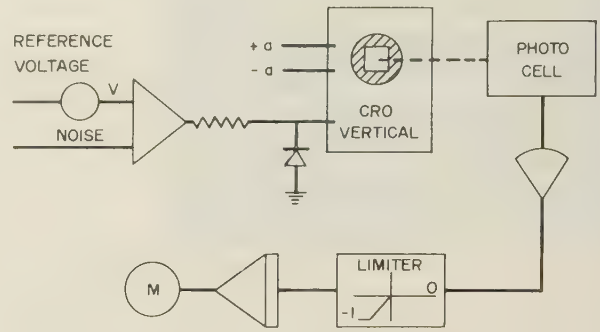


Fig. 1—Circuit for the determination of cumulative probability distribution.

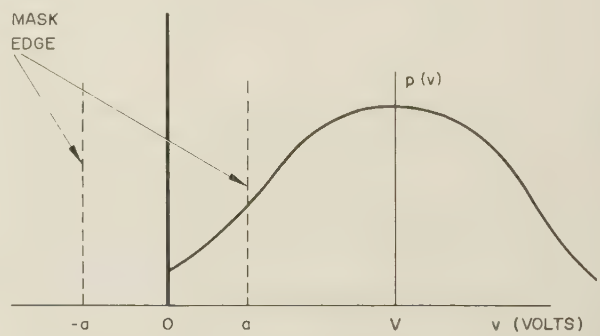


Fig. 2—Probability density of the vertical input.

$$\int_{-\infty}^{a} p(v-V)dv = \int_{-\infty}^{a-V} p(v)dv = P(a-V). \qquad (1)$$

The diode action of the circuit is essentially perfect if the region between $\pm\frac{1}{2}$ volt and, of course, very high voltages are excluded. This is accomplished if a is approximately 1 volt or greater.

If a narrow slot is cut in the mask, it is possible to count the average number of crossings per second at a given level for an arbitrary noise. An analytical solution has been given by Rice:[2]

expected number of zeros per second

$$= \frac{1}{\pi}\left[-\frac{\phi''(0)}{\phi(0)}\right]^{1/2}, \qquad (2)$$

where $\phi(\tau)$ is the autocorrelation function of the noise. An experimental determination of this quantity can be useful in a problem to be discussed later.

A mechanization of the method for the automatic tracking problem, discussed by Laning and Battin,[3] is

* Manuscript received by the PGEC, March 26, 1957. Part of this work was done under Contract AF 33(600)-23101.

† Hughes Aircraft Co., Culver City, Calif.

[1] P. M. Morse and G. E. Kimball, "Methods of Operations Research," John Wiley and Sons, Inc., New York, N. Y., p. 122; 1951.

[2] S. O. Rice, "Mathematical theory of random noise," *Bell Sys. Tech. J.*, vol. 24, p. 51; January, 1945.

[3] J. H. Laning and R. H. Battin, "Random Processes in Automatic Control," McGraw-Hill Book Co., Inc., New York, N. Y., p. 163 ff.; 1956.

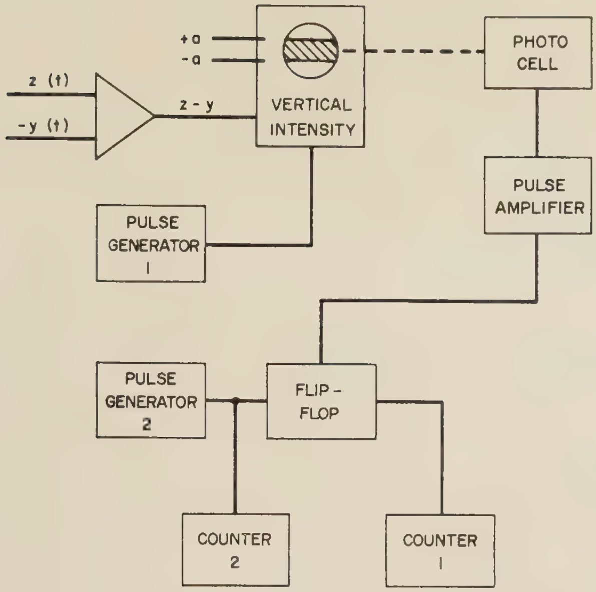


Fig. 3—Mechanization of the automatic tracking problem.

shown in Fig. 3. A signal $y(t)$ and noise $x(t)$ are applied to a servo with resultant output $z(t)$. Tracking will be assumed lost if $z(t)-y(t)$ exceeds a bound a. The system performance may be described by the probability that it loses track in a specified interval Δt.

The mask in this case covers the region between $\pm a$ so that the photocell will record only excursions outside the bound. Pulse generator 1 has a repetition frequency high enough to closely sample the vertical input. Pulse generator 2 defines the interval Δt by resetting the flip-flop at the beginning of each interval so that only one pulse per interval can be counted by counter 1. The ratio of the two counts is thus the probability of at least one excursion beyond $\pm a$ in the interval Δt.

These examples illustrate types of problems to which the method is applicable. Others have been discussed by Favreau, Low, and Pfeffer,[4] who give solutions by the more conventional analog methods.

Some time-varying problems may be conveniently handled by utilizing the freedom of the two-dimensional presentation. For example, if a linear sweep is applied to the horizontal input in the tracking example, the mask may be cut to make the bound a an arbitrary function of time during the interval.

Design Considerations

Instrumentation of the method for a particular type of problem generally involves compromises between circuit technique, accuracy, and computation rate. Limitations inherent in circuit technique may occur in the operational amplifier bandwidth, oscilloscope screen response, counting circuits, etc. These factors define or restrict the noise spectra which may be used, in a manner depending upon the mechanization.

Two general classes of problems are encountered. The noise characteristics may be specified by the problem, or the noise may merely furnish random samples. In the former case, it is necessary only to choose the proper time-frequency scale to provide minimum computation time consistent with the circuit limitations. In the latter case, there is considerable freedom in the choice of the spectrum, and it may be necessary to examine the possibilities for optimum accuracy and computation rate.

Since an error analysis of the solution is generally much more difficult than the problem itself, it is necessary to use results obtained for simple examples of the problem type as a guide.

A representative probability of hit problem has two independent normally distributed aim dispersions with equal variance ϕ_0, directed at the center of a square target with side $2a$. The analytical solution is simply

$$P_h = \frac{1}{2\pi\phi_0}\left[\int_{-a}^{a} e^{-x^2/2\phi_0}dx\right]^2 \qquad (3)$$

and tabulations of the integral are readily available. To mechanize this problem, the aim dispersions are simulated with Gaussian noise generators having an autocorrelation function $\phi(\tau)$ applied to the deflection plates and a mask with a properly scaled rectangular cutout placed over the screen. The photocell output voltage is adjusted to unity if the spot is inside the cutout and zero outside. The output is integrated for a period T and the result divided by T to obtain a mean voltage equal to the sample probability of hit, P_{hs}. Considering only the errors resulting from the noise statistics the expected mean square fluctuation in P_{hs} is given by[5]

$$\sigma_P{}^2 = \frac{2}{T}\int_0^T \left(1 - \frac{x}{T}\right)[\theta(x) - P_h{}^2]dx, \qquad (4)$$

where $\theta(\tau)$ is the autocorrelation function of the photocell output. For large T,

$$\sigma_P{}^2 \cong \frac{W(0+)}{2T}, \qquad (5)$$

where $W(f)$ is the power spectral density corresponding to $\theta(\tau)$.

$$W(f) = 4\int_0^\infty \theta(\tau)\cos 2\pi f\tau d\tau. \qquad (6)$$

In order to increase confidence in the sample probability of hit, it is necessary to increase the integration time for a given noise spectrum.

If the oscilloscope intensity is pulsed with period T_0 and the integration performed by counting these samples of the photocell output, the dispersion will be increased.[6] We have

[4] R. R. Favreau, H. Low, and I. Pfeffer, "Evaluation of complex statistical functions by an analog computer," 1956 IRE Convention Record, part IV, pp. 31–37.

[5] W. B. Davenport, Jr., R. A. Johnson, and D. Middleton, "Statistical errors in measurements on random time functions," *J. Appl. Phys.*, vol. 23, p. 377; April, 1952.

[6] *Ibid.*, p. 380.

$$P_{hs} = \frac{N_h}{N}, \tag{7}$$

$$\sigma_P{}^2 = \frac{\theta(0) - P_h{}^2}{N} + \frac{2}{N^2} \sum_{n=1}^{N-1} (N - n)[\theta(nT_0) - P_h{}^2], \tag{8}$$

where N_h = number of hits, N = total number of pulses. To increase the accuracy it is thus desirable to use as many samples as possible which, however, also increases the computation time, NT_0.

Assuming that Gaussian noise characterized by $\phi(\tau)$ is passed through a nonlinear filter $F(x)$, the correlation function of the output may be obtained by a method given by Laning and Battin[7] for the solution of the similar problem of noise through a limiter.

In general,

$$\theta(\tau) = \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv q(u) q(v) e^{-1/2(\phi_0 u^2 + 2\phi uv + \phi_0 v^2)} \tag{9}$$

where

$$F(x) = \int_{-\infty}^{\infty} q(u) e^{ixu} du. \tag{10}$$

For independent Gaussian sources applied to a two-dimensional filter

$$F(x, y) = \int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} du q(s) r(u) e^{i(xs+yu)}, \tag{11}$$

we have

$$\theta(\tau) = \int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv q(s) q(t) r(u) r(v) \cdot e^{-1/2(\phi_0 s^2 + 2\phi st + \phi_0 t^2) - 1/2(\phi_0 u^2 + 2\phi uv + \phi_0 v^2)}. \tag{12}$$

The oscilloscope, mask, and photocell have a filter characteristic

$$F(x, y) = \begin{cases} 1 & |x| < a \text{ and } |y| < a \\ 0 & |x| > a \text{ or } |y| > a \end{cases} \tag{13}$$

which has an integral representation

$$F(x, y) = \frac{1}{\pi^2} \int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} du \frac{\sin as}{s} \frac{\sin au}{u} e^{i(xs+yu)}. \tag{14}$$

Thus,

$$q(u) = r(u) = \frac{\sin au}{\pi u}, \tag{15}$$

$$\theta(\tau) = \frac{1}{\pi^4} \int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv \frac{\sin as}{s} \frac{\sin at}{t} \frac{\sin au}{u} \frac{\sin av}{v} \cdot e^{-1/2(\phi_0 s^2 + 2\phi st + \phi_0 t^2) - 1/2(\phi_0 u^2 + 2\phi uv + \phi_0 v^2)}. \tag{16}$$

Using a series expansion for the exponential, this may be written

$$\theta(\tau) = \left\{ \sum_{n=0}^{\infty} \frac{(-1)^n \phi^n}{n!} \left[\frac{1}{\pi} \int_{-\infty}^{\infty} u^{n-1} \sin au e^{-\phi_0 u^2/2} du \right]^2 \right\}^2. \tag{17}$$

Noting that the terms for $n = 1, 3, 5, \cdots$, vanish, we have

$$\theta(\tau) = \left\{ \sum_{n=0}^{\infty} \frac{\phi^{2n}}{(2n)!} \left[\frac{2}{\pi} \int_{0}^{\infty} u^{2n-1} \sin au e^{-\phi_0 u^2/2} du \right]^2 \right\}^2. \tag{18}$$

Following the method of the above reference this expression may be reduced to

$$\theta(\tau) = [P_h + S(\rho)]^2 \tag{19}$$

where

$$\frac{dS}{d\rho} = \frac{1}{\pi\sqrt{1 - \rho^2}} \left[e^{-x^2/(1+\rho)} - e^{-x^2/(1-\rho)} \right] \tag{20}$$

with

$$\rho(\tau) = \frac{\phi(\tau)}{\phi_0} \tag{21}$$

$$x = \frac{a}{\sqrt{\phi_0}}. \tag{22}$$

Difficulties in numerical integration in the vicinity $\rho = 1$ may be avoided by an independent calculation from the density function of the photocell output which may be written

$$p(v) = (1 - P_h)\delta(v) + P_h \delta(v - 1) \tag{23}$$

from which

$$\theta(0) = \int_{-\infty}^{\infty} v^2 p(v) dv = P_h = [P_h + S(1)]^2 \tag{24}$$

thus

$$S(1) = \sqrt{P_h} - P_h. \tag{25}$$

Fig. 4 gives θ as a function of $\rho(\tau)$ for several values of a^2/ϕ_0. Using these curves it is possible to obtain the autocorrelation function of the photocell output for any Gaussian noise source. The expected mean-square error in the sample probability of hit is then given by (4), (5), or (8).

Examples of Design

The following examples serve to illustrate the application of the results of the preceding section to a problem where amplifier bandwidth limits performance and to one where the oscilloscope screen response is the limiting factor.

Simple Probability of Hit

A mechanization of the problem discussed with regard to statistical errors is shown in Fig. 5. The oscilloscope screen would limit the accuracy if the response time to the limiter output of -1 volt becomes com-

[7] Laning and Battin, *op. cit.*, p. 169.

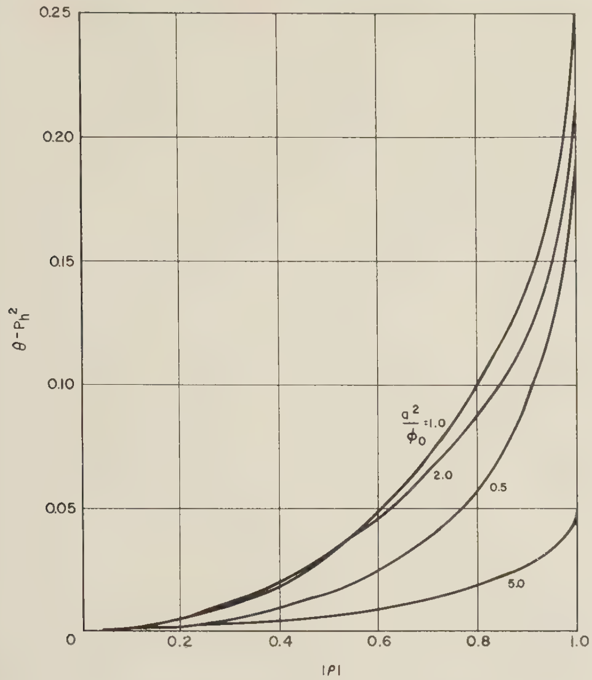


Fig. 4—Autocorrelation function of the photocell output.

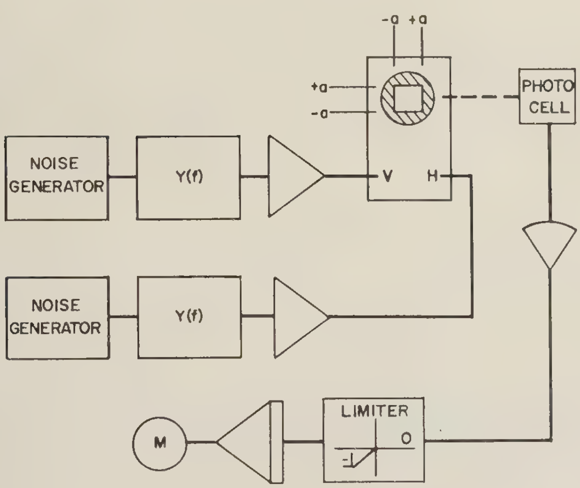


Fig. 5—Circuit for the determination of probability of hit.

parable with the average length of time the spot spends within the cutout. This results in a rounding of the normally rectangular output voltage to the integrator which depends on the noise bandwidth, and its variance with respect to the target dimensions. The amplification between the photomultiplier and the limiter effectively improves the screen response but is limited by noise from the photo tube. The effects of these circuit factors cannot be described in a simple manner.

Another limitation arises from the response of the operational amplifiers. In this simple problem the amplifier response could normally be taken into account directly. However, if several sources of dispersion affecting both channels are to be combined in several amplifiers it becomes necessary to insure that the rms noise voltages are not affected significantly by the amplifiers.

A possible criterion which eliminates this source of error is to restrict the noise bandwidth such that only one per cent of the total noise power is contained in frequency components above a maximum frequency f_m, determined by the operational amplifiers. If the basic noise source has too wide a power spectrum its output must be filtered until it is consistent with the above requirement.

The characteristics of three of the many possible filters, $Y(f)$, are given in Table I which compares single, double, and triple isolated, RC, low-pass networks with white noise input. The expected variance, σ_P^2, of the measured P_{hs} was determined for large T using Fig. 4 and (5) for $\phi_0 = a^2$.

TABLE I

COMPUTATION EFFICIENCIES OF RC LOW-PASS FILTERS OF MAXIMUM FREQUENCY f_m

Filter	$\lvert Y(f) \rvert^2$	$\phi(\tau)$	$f_{m/fn}$	$f_m \sigma_P^2 T$
One RC	$\frac{f_1^2}{f^2+f_1^2}$	$\phi_{01}e^{-\omega_1\tau}$	$f_{m/f_1}=63.7$	1.63
Two RC	$\frac{f_2^4}{(f^2+f_2^2)^2}$	$\phi_{02}e^{-\omega_2\tau}(\omega_2\tau+1)$	$f_{m/f_2}=3.35$	0.251
Three RC	$\frac{f_3^6}{(f^2+f_3^2)^3}$	$\phi_{03}e^{\ \omega_3\tau}\left(\frac{\omega_3^2\tau^2}{3}+\omega_3\tau+1\right)$	$f_{m/f3}=1.80$	0.152

$$\omega_n = 2\pi f_n = \frac{1}{RC}, \quad \phi_0 = a^2.$$

TABLE II

MINIMUM PULSE REPETITION PERIOD FOR RC FILTERS

Filter	$\phi''(0)/\phi_0$	$\frac{v_m}{r}\tau_0$
Two RC	$-\omega_2^2$	23.3
Three RC	$-\frac{\omega_3^2}{3}$	17.1

It is apparent from Table II that a sharp frequency cutoff characteristic is desirable if the maximum frequency criterion is used since, given f_m and σ_P^2, the triple RC filter allows over ten times as high a computation rate as the single RC filter.

Airborne Rocket Fire Control System

As an example of a problem to which this method may be profitably applied and which illustrates the manner in which screen response may limit the accuracy and computation rate, consider the case of a salvo of unguided, contact-fuzed rockets, launched, equally spaced in an interval which is short compared to the response time of the control system of the launch-

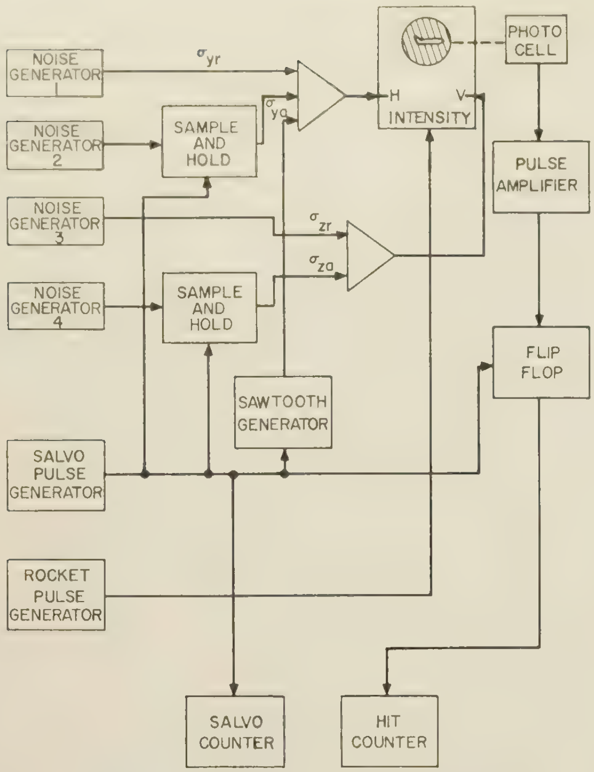


Fig. 6—General arrangement for determining rocket probability of hit.

ing aircraft. The motion of the aim point along the target during the salvo is then a function only of the relative velocity and attack geometry. It will be assumed that one hit in the given vulnerable area is sufficient for a kill so that only the first hit in a salvo should be counted. Random errors in the rocket trajectories due to aiming errors and rocket dispersion must be considered separately, but each may be assumed independent and normal in two dimensions. Four independent noise sources are thus required.

A block diagram of the mechanization is shown in Fig. 6, where σ_{ya}, σ_{yr}, σ_{za}, and σ_{zr} are the azimuth and elevation aim and rocket dispersions, respectively.

The salvo begins with a pulse from the salvo pulse generator which selects a new aim point from noise generators 2 and 4, triggers the sawtooth generator simulating the linear aim point motion, and resets a flip-flop to respond to a possible hit recorded by the photocell. The output of the flip-flop, which can respond only once during the salvo, is counted by the "hit" counter. The number of rockets in a salvo is determined by the relative frequencies of the "salvo" pulse generator and the "rocket" pulse generator which is connected to the intensity grid of the oscilloscope.

A basic requirement in this simulation is that the individual rocket dispersions be uncorrelated. In general, the noise bandwidth must be increased if this requirement is to be met while increasing the rocket pulse rate and thus the computation rate.

There is, however, an associated increase in the velocity of the spot on the oscilloscope screen. This elongates the scintillations and effectively blurs the sharp boundaries of the mask. The length of a trace is simply $v\Delta t$ where v is the average spot velocity during a pulse length Δt. The minimum pulse length is generally determined by the response of the screen and is of the order of $\frac{1}{2}$ μsec depending on the operating conditions of the tube and the sensitivity of the photomultiplier.

The amount of degradation in accuracy resulting from this effect is not easily estimated in general since it depends on the shape of the mask and the portion of a trace required to give a count. If the trace lengths are not significantly larger than a stationary spot, the accuracy will not be degraded. If such a requirement results in too long a computation time, as determined from the following analysis, the effect of increased trace length can best be determined by observing the effect of increasing pulse lengths on a measured probability of hit.

If a trace length $v_m\Delta t$ is the maximum consistent with the resolution of the mask, the noise spectrum must be chosen such that the spot will appear on the screen with a velocity greater than v_m only a small percentage of the time.

An estimate of this requirement may be obtained by assuming a screen radius r with independent Gaussian sources of equal standard deviation σ (inches) applied in the vertical and horizontal channels and centered on the screen. The probability of a trace appearing on the screen is

$$1 - e^{-r^2/2\sigma^2}$$

and the probability of the spot velocity being greater than v_m is

$$1 - \left[1 - e^{-v_m^2/2\sigma_v^2}\right] = e^{-v_m^2/2\sigma_v^2} \tag{26}$$

where σ_v is the standard deviation of the velocity distribution. Since the amplitude and velocity distributions are independent, the probability of a trace longer than $v_m\Delta t$ appearing is

$$P = e^{-v_m^2/2\sigma_v^2}\left[1 - e^{-r^2/2\sigma^2}\right]. \tag{27}$$

Let

$$k^2 = \frac{\sigma_v^2}{\sigma^2} = -\frac{\phi''(0)}{\phi(0)} \tag{28}$$

where $\phi(\tau)$ is the autocorrelation function of the noise and it is assumed that $\int_0^\infty f^2W(f)df$ exists. Then

$$P = e^{-v_m^2/2k^2\sigma^2}\left[1 - e^{-r^2/2\sigma^2}\right]. \tag{29}$$

In order to insure meeting the criterion, P may be determined for the worst σ. Maximizing this function one finds

$$P_{\max} = \frac{1}{1+\mu^2}\left[\frac{\mu^2}{1+\mu^2}\right]^{\mu^2} \tag{30}$$

where

$$\mu = \frac{v_m}{rk}. \tag{31}$$

This function is plotted in Fig. 7.

If a correlation of 0.1 for successive rockets is acceptable and only one per cent of the traces have a velocity greater than v_m, the allowable rocket pulse repetition frequency $1/\tau_0$ may be computed from the correlation function of the noise sources and the value $k = v_m/6r$ obtained from Fig. 7.

The two RC filters previously considered for which the rms velocity exists are compared under these conditions in Table II. Here it is also generally desirable to use a sharp cutoff frequency characteristic in the noise filter.

When the maximum rocket pulse rate has been established, the accuracy depends on the sample size, *i.e.*, the number of salvos fired. The photocell output is subject to additional filtering by the "at least one" circuit. However, the resulting probability distribution is that of (23) and the dispersion of the result may be obtained as a special case of (8). Since the results of the salvo firings are uncorrelated,

$$\sigma_P^2 = \frac{P_h(1 - P_h)}{N} \tag{32}$$

where N is now the number of salvos. This result may also be obtained from a consideration of the binomial distribution of which this is an example.

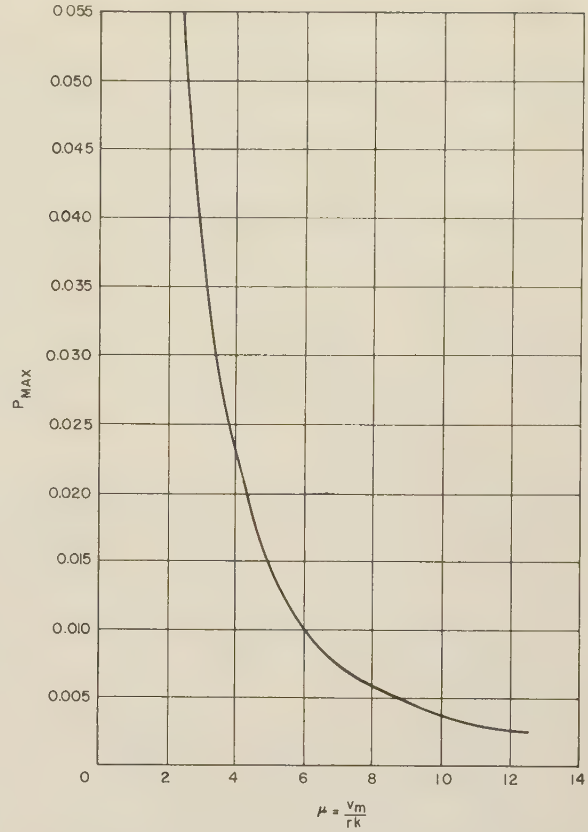


Fig. 7—Maximum probability of a trace appearing inside a radius r and having a velocity $> v_m$.

An Experiment in Musical Composition*

F. P. BROOKS, JR.†, A. L. HOPKINS, JR.‡, P. G. NEUMANN‡, AND W. V. WRIGHT‡

Summary—**The high-order probabilities of element sequences can be determined from a sample of linear structures and can be used for synthesis of new structures. From theoretical considerations one can identify the qualitative conditions for satisfactory output. The theoretical concepts can be tested and quantitative parameters determined by experiment. Such an experiment has been performed by analyzing written music and by testing the analysis through the synthesis of new musical compositions, using a digital computer.**

A sample of 37 melodies was analyzed for the probabilities of the elements, element pairs (digrams), trigrams, and so on to the eighth order. The tables derived were used for the synthesis of original melodies by a random process. The theory and the experimental verification are considered in detail. The experimental results presented include comparative statistics of the successful syntheses using each of the eight orders of analysis, examples of melodies generated by low, medium, and high-order synthesis, and confirmation of degeneracy and other effects predicted by the theory.

* Manuscript received by the PGEC, April 20, 1957.

† Product Dev. Lab., International Business Machines Corp., Poughkeepsie, N. Y. Formerly with The Computation Lab., Harvard Univ., Cambridge, Mass.

‡ The Computation Lab., Harvard Univ., Cambridge, Mass.

Introduction

HUMAN thought processes have at least two radically different components, often identified as induction and deduction. Digital computers are readily programmed to perform deductive "reasoning," but their ability to draw generalizations from special cases is extremely limited. Many interesting computer experiments in game playing, "learning," theorem proving, etc., have been aimed at discovering methods of simulating rudimentary inductive processes with a computing machine.[1,2]

[1] C. E. Shannon, "Computers and automata," Proc. IRE, vol. 41, pp. 1234–1241; October, 1953.

[2] A. G. Oettinger, "Programming a digital computer to learn," *Phil. Mag.*, vol. 7, pp. 1243–1263; December, 1952.

There are often demands for inductive reasoning where the results of the generalization process do not need to be stated explicitly as rules but only need be in a form suitable for subsequent deductive reasoning. This is the case whenever one attempts to synthesize structures of a certain class, as in the creation of synthetic linguistic utterances or synthetic musical compositions. The simplest way to perform such tasks is for a human to analyze some sample of the type of structure desired, draw up some explicit rules and constraints, and allow the machine to operate deductively, although perhaps at random, in the synthesis of new structures. Some workers have performed computer-implemented musical composition in this manner.

It is of considerably more interest to attempt to synthesize musical compositions by having the machine inductively analyze a sample of acceptable compositions and, using its conclusions, deductively synthesize new but original compositions. Such an induction can be performed by determining the probabilities of note sequences. Several theoretical aspects of such a process deserve examination.

Theory of Analysis-Synthesis

The derivation of sufficient information from a sample to permit subsequent deductive synthesis of new and original members of the same class of structure depends not only upon the sophistication of the analysis, but also upon the characteristics of the sample analyzed. Suppose one undertakes to analyze a small number of tunes and to use the results in a random process for synthesizing new tunes. One can anticipate three causes of difficulty which can be visualized with the aid of the diagram in Fig. 1. This diagram shows that the

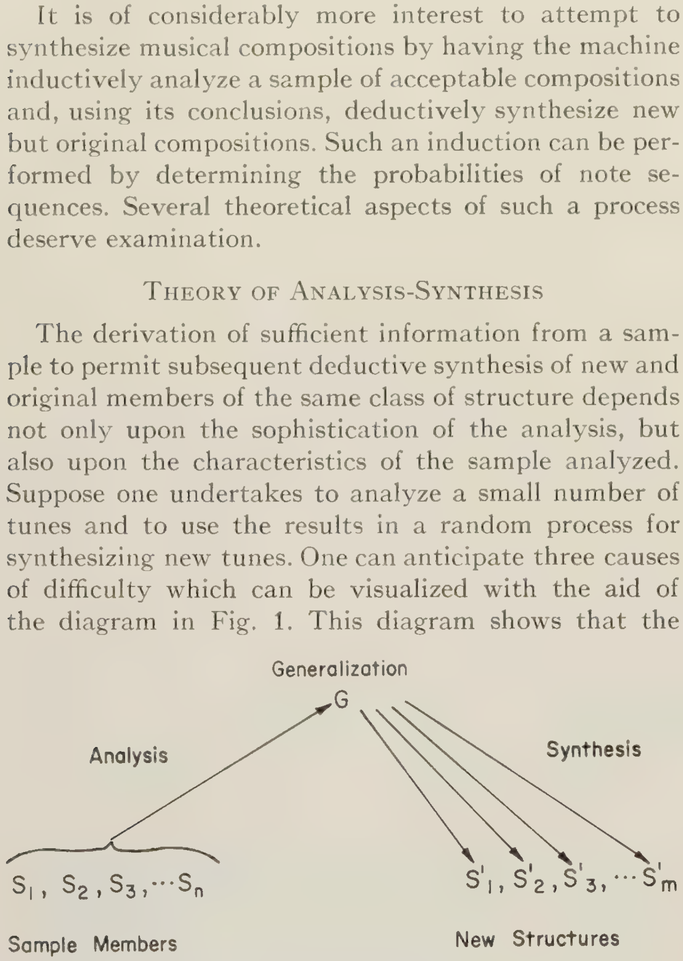


Fig. 1.

process of analysis of several members of a sample $S_1, S_2, \cdots, S_n$, belonging to some common class of structure, yields a generalization G. From the generalization, a synthesis is used to derive one or more new structures S', which one hopes will be of the same class of structures as the original sample. The first difficulty is that overly naive analysis may yield a generalization so loose that the resulting S' does not belong to the same class of structure as the sample members. For example, synthetic note sequences may not belong to the class of acceptable melodies from which they were derived. The second difficulty is that the sample may be so small that no generalization may be drawn no matter how sophisticatedly an analysis is performed. The third difficulty is that the sample members may be so alike that when the generalization is formed, it is impossible to create a new structure of the same class which is not identical to one of the structures of the sample.

It appears worthwhile to perform some experiments to verify the existence of these three effects and to learn something about the basic nature of the analysis and synthesis process. For this purpose, two structure classes of interest are the class of linguistic utterances acceptable to human beings as "meaningful" and the class of musical compositions acceptable to human beings as something a human might have created. The synthetic linguistic utterance problem (which includes the well-known sonnet-writing problem) is more general, more interesting, and more difficult than the musical composition problem. The musical composition problem, interesting in its own right, also serves as a small-scale model of the linguistic utterance problem. For these reasons the experiments described herein were undertaken on the application of the analysis-synthesis process to the musical composition model.

The analysis of any structural system consists of the determination of a set of basic elements and the determination of the combinational relationships among the elements. For the present, we consider that only the second step is sufficiently difficult mechanically to be worth computer mechanization and sufficiently simple conceptually to permit computer mechanization.

The constraints of combination of an element set can be stated in several different ways. The first of these is the *explicit* statement of rules as to what combinations may or may not occur. For example, the explicit method of representing linguistic constraints is illustrated by the schoolboy's grammar and spelling books. Here one finds several specific examples of a connective pattern grouped together with the generalization from these to a rule: "*i* before *e*, except after *c*, or when sounded as *a*, as in *neighbor* or *weigh*." From the set of rules on spelling, the explicit method extends to the formulation of rules for the combination of larger elements in formal grammar, syntax, and rhetoric. The expression of all the combinational constraints in a structure by such a method may grow quite complex with many levels of rules, sub-rules, exception systems, and specific exceptions. Examination of any complete grammar reveals this complexity at which the "*i* before *e*" example only hints.

A second approach is the *exhaustive* one in which all the combinational constraints are listed explicitly. A class can always be completely described by listing its members, and this description is often more useful, more revealing, and less confusing than any set of class characteristics. We do, in fact, learn the constraints of combination for *i* and *e* by long and painful use of spelling book and dictionary which list all the extant combinations. By this means, our description of constraints extends to *seize* and *financier* without any special treatment. While the exhaustive approach has the advantage

of conceptual simplicity and guarantees a complete description of any structure, it is tedious and burdensome to apply to any system as complex as language or music.

Both the explicit and the exhaustive methods of constraint description are incomplete; they only indicate which combinations may occur without describing the relative frequency of occurrence. The mere statement that t and e can each follow an initial p is incomplete without the information that *pt-* is much less common than *pe-*. This leads to a third method of constraint description, the probabilistic statement.

The probabilistic description shares the advantages of the exhaustive description; it is simple in concept and uniform in application. Ease and certainty of determining combinational constraints in analysis and following them in synthesis are independent of the complexity or obscurity of the constraints themselves. For computer application these are cardinal virtues, and the great magnitude of the task of formulating and applying such a description is a less important problem. The probabilistic method has a further advantage over other methods: the precision and validity of the description are little affected by the inclusion or exclusion of sample members beyond a sample of a certain size. With both the explicit and the exhaustive methods, the existence of an exception not included in the sample analyzed would compromise the accuracy of the analysis. Finally, although analyses for some purposes are required to yield an explicit statement of the common properties of a sample, this is not required for the analysis-synthesis process shown in Fig. 1.

For these reasons, the probabilistic method was chosen for the analysis-synthesis experiments described below. Analysis by probabilities of combinations may be taken to any order. One may use simple probabilities of occurrence, simple transitional probabilities in a sequential, linear structure, or conditional probabilities depending upon any number of preceding or adjacent elements.

The concept of using probabilistic analysis as part of an analysis-synthesis procedure is by no means new. Shannon[3] proposed a simple and elegant method of performing an implicit analysis and synthesis of linguistic utterances. Pinkerton[4] and F. and C. Altneave[5] have used simple transitional probabilities for the analysis and synthesis of melodies.

The use of probabilistic descriptions of several orders, however, permits one to apply successively more sophisticated analytic procedures to a given sample in an attempt to determine the existence and behavior of the three hindering effects set forth earlier. This was the primary purpose of the experiments described below. It was an incidental but fervent hope that the sample selected would be sufficiently large and rich to permit some of the more sophisticated analysis-synthesis procedures to yield musical compositions that were both original and musically acceptable.

For structures that are linear or one-dimensional, higher-order transitional probabilities are known as *Markoff chains*, and we shall refer to their determination as a *Markoff analysis* and their use in the construction of new structures as a *Markoff synthesis*. Let us examine in more detail the analytic and synthetic methods used.

Method

A Markoff analysis of order m is the process of determining certain joint probabilities of the occurrences of elements in a sample. Thus, within the sample, a given sequence of $(m-1)$ elements may be followed by any one of several elements. We wish to find the relative frequency of each of these elements after the given $(m-1)$ element sequence. A sequence of m elements will be called an *m gram*. A sample will ordinarily be made up of several independent sequences of elements called *utterances*. A step in a Markoff synthesis of order m is performed by choosing an element to follow a given element sequence of length $(m-1)$ in such a way that, if this process were to be repeated a large number of times, the relative frequencies of the chosen elements would match those found in the sample. Since it is desired that the synthesis be random within these restrictions, the choice is made according to the magnitude of a random or pseudo-random number. It is sufficient for this purpose, then, to construct for each element sequence of length $(m-1)$ a table of cumulative probabilities which allots to each mth element a segment of the range of numbers between zero and one, the size of the segment being the desired relative frequency of the note. Then a random number between zero and one falls into the range allotted to an element with a probability equal to the relative frequency of the element. This procedure is, in effect, a mapping of a uniform distribution of random numbers into the nonuniform distribution of elements following a given sequence of $(m-1)$ elements. Consider, for the sake of illustration, the following hypothetical fourth-order analysis and synthesis of musical text. Suppose that the occurrences of a three-note sequence $C-E-G$ in the sample are as shown below:

$$C-E-G-A$$
$$C-E-G-A$$
$$C-E-G-C$$
$$C-E-G-E$$
$$C-E-G-E$$
$$C-E-G-G$$
$$C-E-G-G$$
$$C-E-G-G$$
$$C-E-G-G$$
$$C-E-G-G.$$

[3] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication," Univ. of Illinois Press, Urbana, Ill., pp. 11–15; 1949.

[4] R. C. Pinkerton, "Information theory and melody," *Sci. Amer.*, vol. 194, p. 77; February, 1956.

[5] F. and C. Altneave, unpublished study described by H. Quastler, "London Symposium on Information Theory," Butterworth, Ltd., London, England, pp. 168–169; September, 1955.

The relative frequencies of notes following the sequence $C-E-G$ are:

A	$2/10 = 0.2$
C	$1/10 = 0.1$
E	$2/10 = 0.2$
G	$5/10 = 0.5.$

Synthesis may then be carried out by means of the table of cumulative probabilities which, for any argument, x, between zero and one, will determine the note to be selected. One of the several possible arrangements of such a table is:

A	$0 \leq x < 0.2$
C	$0.2 \leq x < 0.3$
E	$0.3 \leq x < 0.5$
G	$0.5 \leq x < 1.0.$

If the random numbers x are uniformly distributed between zero and one, the probability of choosing a C is one tenth; and in the course of synthesis, the sequence $C-E-G$ will be followed by C with expectation of occurrence of one tenth. Clearly, the conditional probabilities found by Markoff analysis can be approximated to any desired accuracy by this method.

The three hindering effects considered above can be understood more precisely in terms of the Markoff analysis and synthesis method. The first effect, overly naive analysis, will show itself when only low-order transition probabilities are used, so that the structure of the synthetic composition is insufficiently constrained. This permits the occurrence of note sequences that are not acceptable as legitimate musical structures.

The second effect, insufficient sample size, shows itself at some order m of analysis and synthesis where each distinct m gram occurs only once. From that order upward, further analysis is useless since the higher-order transition probabilities are completely determined by the analysis to that point. There is no more information to be extracted from the sample. Therefore, if the synthesis at this order fails to yield acceptable and original utterances, nothing can be done, for the sample size is so small that further sophistication of the analysis yields no benefits.

The third effect, insufficient sample diversity, will always show itself before or at the same time as the second effect, and so is of much greater practical importance. Suppose the digrams occurring in the sample have been found, and a trigram analysis is performed yielding one distinct trigram for each distinct digram. Even though there may be many occurrences of each digram and thus of its associated trigrams, an utterance is completely specified by giving the initial digram, for it will specify a trigram whose final two elements uniquely specify another, etc. Further sophistication of the analysis is fruitless beyond this point; all the information in the sample has been extracted. If the analysis to this order is insufficient to yield acceptable and original utterances, nothing can be done; the sample is too redundant to yield more information.

From the theoretical examination of the analysis-synthesis problem, one can see that experiments using Markoff analysis and synthesis will yield any of three results. If too elementary or low-order analysis is used, the results will not resemble the sample members closely enough to be recognized as members of the same class. If too high an order of analysis is attempted for a given sample size and diversity, the synthesized results will degenerate; that is, they will duplicate sample members. Or, if the sample is sufficiently large and diverse, there will be some orders of analysis for which the results are original and still recognizable as members of the class.

For the sample of music selected, each of the three results was found for some order of the Markoff analysis-synthesis.

The Experiments

As indicated above, both linguistic utterances and musical compositions are interesting structure classes upon which to perform experiments in Markoff analysis and synthesis; the musical structures were chosen because of the greater simplicity. The selection of music did introduce problems, however. Music has many dimensions, such as pitch, meter, rhythm, key, harmony, dynamics, and quality. Some selection must be made as to which to include as experimental variables and which to ignore. Since quality is not expressed in written music, and difference in key can be grossly compensated by transposition, these were not treated as variables. Harmony introduces a vertical as well as a horizontal structure to the text, and the need for simplicity dictated that the present experiments be confined to the horizontal structure of a single line. The metrical structure was fixed as an experimental variable by using a sample all of whose members had a fixed metrical structure.

After written music was chosen as the structure class to be used in the experiments, a sample was assembled. It consisted of 37 common meter hymn tunes, a choice determined largely by the decision to work within one metrical structure. The common meter hymn is perhaps the most widely used rigid metrical structure and one of the simplest. In fact, several variants are found even within this structure, and the sample was confined to the most common single variant. The common meter hymn tune has the additional advantage of providing a fairly large collection of compositions from different composers and centuries of origin. The hymns in the sample all begin on the last beat of a four-beat measure, and none have any notes shorter than an eighth note.

Several analyses of the sample and syntheses of new hymn melodies were performed with a large-scale com-

puter. The machine has random-access storage of 230 sixteen-decimal digit numbers and separate magnetic drum storage of 4000 numbers and 10,000 instructions. It is synchronous in operation with most instructions requiring 1.3 μsec.

For computer manipulation it was necessary to encode the notes as numbers. A range of four octaves in the chromatic scale was selected, and each tone was assigned a two-digit number. Different time values were represented by dividing the whole hymn into 64 eighth-note cells. The content of each cell was either a tone struck in that cell or one held over from the preceding cell. The even integers from 02 to 98 were used to represent a struck tone while the corresponding odd integers from 03 to 99 were used to represent a continued tone. Henceforth, each of the eighth-note cells will be referred to as a note regardless of whether it represents a struck or a held note. In order to provide a common basis for the analysis, each hymn was transposed upwards to the nearest key of *C*.

As each storage location was capable of holding sixteen decimal digits and hence eight notes, it was convenient to carry out analyses up to the eighth order. As it turned out, this order was the one at which redundancy-caused degeneracy became noticeable.

The first step in the analysis was to isolate all of the eight-note sequences or octograms occurring in the sample, and then sort them into numerical order.[6]

This sort placed all of the octograms in order within their initial heptagram, all heptagrams in order within their initial hexagrams, and so on. Hence, the sorted sequences of lower orders were readily obtainable from the sorted octograms by shifting all the octograms the appropriate number of places to the right, thus eliminating the extraneous notes.

To complete the analysis, cumulative relative frequency tables of the type previously discussed were formed for each value of m for all of the $(m-1)$-note sequences present in the sample. The relative frequency for each distinct m-gram was calculated directly from the sorted m-grams by counting the number of occurrences of the m-gram. This number; *i.e.*, the number of times the mth note followed the initial sequence of $(m-1)$ notes, was divided by the total number of occurrences of all m-grams having the same initial $(m-1)$-note sequence, thus giving the relative frequency of the mth note with respect to that initial sequence. The entries in the table were calculated by accumulating the relative frequencies pertaining to the same $(m-1)$-note sequence.

As an illustration of this procedure, consider the hypothetical sequence of sorted octograms shown in Table I. The resulting octogram table entries are the

TABLE I

SORTED OCTOGRAMS, ILLUSTRATING THE FORMATION OF THE CUMULATIVE PROBABILITY TABLES

Cell 1 2 3 4 5 6 7 8	Octogram Count	Heptagram Count	Relative Frequency	Cumulative Probability
36 37 26 27 32 33 26 22	1			
*36 37 26 27 32 33 26 22	2		2/8	2/8
36 37 26 27 32 33 26 27	1			
36 37 26 27 32 33 26 27	2			
36 37 26 27 32 33 26 27	3			
36 37 26 27 32 33 26 27	4			
*36 37 26 27 32 33 26 27	5		5/8	7/8
*36 37 26 27 32 33 26 28	1	8	1/8	8/8
36 37 26 27 32 33 32 33	1			
*36 37 26 27 32 33 32 33	2	2	2/2	2/2
*36 37 32 33 32 33 32 33	1	1	1/1	1/1
36 37 32 33 32 33 36 37	1			
*36 37 32 33 32 33 36 37	2		2/3	2/3
*36 37 32 33 32 33 36 42	1	3	1/3	3/3

distinct octograms (denoted with an asterisk) and their associated cumulative probabilities, shown in the last column. The octogram 3637 2627 3233 2622 corresponds to the eight-note sequence $G\bar{G}C\bar{C}A\bar{A}CD$, where a bar over a letter indicates that the note is held.

For each m the number of these tables was equal to the number of distinct $(m-1)$-grams. Table II, next page, gives the total number of m-grams beginning with struck notes, held notes, and initial rests (00). With 64 octograms in each hymn, the 37 hymns yielded 2368 octograms. As seen from the table, only 1701 of these were distinct. Since the number of distinct heptagrams was 1531, there were exactly that many octogram tables having 1701 entries in all. The 4000-word drum storage of the computer thus sufficed for the storage of both the sorted distinct octograms and their associated cumulative probabilities.

THE SYNTHESIS

Syntheses of orders one through eight were then accomplished by the process of Markoff synthesis discussed earlier, using eight digit pseudo-random numbers obtained from an algorithm known to be non-repetitive for the first $(10^8+1)/17$ numbers.[7] The first note of the hymn was found by entering the probability table for the m-grams whose first $(m-1)$ notes were rests and choosing at random among these. The second note was selected by choosing an m-gram beginning with $(m-2)$ rests followed by the first chosen note. The procedure of choosing one m-gram, given the preceding $(m-1)$ notes, was carried on in similar fashion until the 64 notes comprising one hymn had been generated. The encoding of initial rests preceding each hymn thus per-

[6] The sort was performed by a general purpose digital sorting routine that counts digital occurrences in the $(n+1)$st column while sorting on the nth column. Rapid and efficient in use of storage, the routine sorts up to 2000 numbers without requiring any intermediate input or output. The authors are indebted to A. S. Goble III and J. Hines for the programming of the routine.

[7] D. H. Lehmer, "Mathematical Methods in Large-Scale Computing Units," *Proceedings of a Second Symposium on Large-Scale Digital Calculating Machinery (1949)*, Harvard Univ. Press, Cambridge, Mass.; 1951.

TABLE II

Distinct M-Gram Counts, Beginning with Several Different Notes

Initial Note	Order of Analysis m							
	1	2	3	4	5	6	7	8
12 G	1	3	15	25	57 –	66	99 –	102
13 $\bar{G}$	1	12	30	67 –	84	124 –	131	150
16 F	1	6	14	25	43 –	45	63 –	63
17 $\bar{F}$	1	8	14	32 –	33	51 –	51	66
18 E	1	6	19	32	74 –	83	131 –	131
19 $\bar{E}$	1	11	23	68 –	78	130 –	135	156
22 D	1	4	19	30	66 –	69	104 –	109
23 $\bar{D}$	1	12	21	57 –	64	102 –	106	126
26 C	1	6	18	28	65 –	71	111 –	112
27 $\bar{C}$	1	13	25	65 –	72	112 –	113	136
All Struck Notes	18	47	152	219	444 –	479	698 –	705
All Held Notes	18	110	182	428 –	485	717 –	738	869
00 Initial Rest	1	5	10	28	45	70	95	127
Total Distinct m-Grams	37	162	344	675	974	1266	1531	1701

TABLE III

Percentage of Attempts Yielding Acceptable Hymns

Metric Constraint	Order of Analysis m							
	1	2	3	4	5	6	7	8
Even Quarters	100	40	32	10.8	9.2	3	2.5	8.4
Quarters with Various Options		41	26	15	11	2		10
Dotted	100	13.5	5	2	1	0	0	0
Dotted with Various Options	100	17	4	1.5	0	0	0	0
Skeleton	100			5	4	3	2.4	10

TABLE IV

Example of M-Gram Extension Properties

m	Note Brought in	m-Gram, Begun with Struck Note	m-Gram, Begun with Held Note
4	E	$C\ \bar{C}\ D\ \bar{D}$	$\bar{C}\ D\ \bar{D}\ E$
5	E	$C\ \bar{C}\ D\ \bar{D}\ E$	$\bar{C}\ D\ \bar{D}\ E\ \bar{E}$
6	F	$C\ \bar{C}\ D\ \bar{D}\ E\ \bar{E}$	$\bar{C}\ D\ \bar{D}\ E\ \bar{E}\ F$

mitted the synthesis process to operate uniformly, even in starting.

The selection of notes was subject to certain externally applied constraints in addition to those implicit in the frequency tables of the analysis. These explicit constraints, however, were just those permitted by the selection of a sample of uniform metrical structures. In order to force the synthesis to stay within the selected metrical structure, a metric constraint was applied to certain critical notes. For example, the first note of each measure had to be struck rather than held. In addition, the two main phrases both had to end on a dotted half note; hence notes 28 through 32 and 60 through 64 were constrained to be held notes. Each note generated was examined and compared with the metric constraint, if any, for its position in the time basis. If the constraint was not met, the note was rejected and another note was generated by the next random number. If the constraint could not be met after 15 trials, the hymn was discarded and a new one begun. Using these metrical constraints in this manner can be shown to be equivalent to determining and using separate note combination constraint tables for each time point or for subsets of the time points. The rigid metrical structure represented by the sample permitted the separation of metrical constraints from others with a considerable resultant simplification.

An additional constraint was added in the last measure of the hymn. Since after transposition all of the hymns in the sample ended in C above middle C, every generated hymn was required to end the same way. If, after fifteen trials, no final m-gram was chosen which led into a dotted half note C above middle C, the first 57 notes already generated were used to begin another synthesis. This device distorted the absolute size of the yield percentages (number of acceptable hymns completed over the number of starts). It did not, however, affect the faithfulness with which synthesized hymns obeyed all combination probabilities, nor should it have affected the relative sizes of the yield percentages as metric constraint or analysis order changed. Its use permitted the production of a significant number of acceptable tunes within a reasonable time.

Results and Conclusions

The synthesis just discussed was carried out with various metric constraints for all values of m from one to eight. The constraints were of three types. One of these types was based on even quarter notes throughout the hymn, except during the phrase ends. A second was based on a dotted rhythm in the second and sixth measures. Both of these constraints were varied by the introduction of optional rhythms. The third type was a skeletal constraint requiring only that the first and seventh cells in each measure contain struck notes and that the two main phrases end as usual, leaving the remaining rhythmic structure optional.

In all, over 600 complete hymns were synthesized in some 6000 starts. In cases where the yield was a very low percentage of the number started, more attempts were made than in cases where the yield was high. A summary of the results is shown in Table III, which gives the yield for each order m under the various constraints.

The yield is the percentage of the completed hymns out of the total number started. In particular, the similarities between the yields for orders four and five and between the yields for orders six and seven are worthy of comment since they resemble those between the distinct m gram counts in Table II. As predicted from theoretical considerations, the transition from an odd m to the next higher (even) m introduces more implicit constraints than does going from an even m to the next higher (odd) m because of the basic quarter-note structure. Thus,

TABLE V

METRIC CONSTRAINTS OF EXAMPLES 1–5

Ex. 1	X	—	X	—	O	X	X	—	X	X	X	—	X	—	X	—	X	—	X	X	X	X	X	—	X	X	X	—	—	—	—	—
Ex. 2	X	—	X	—	—	X	X	—	X	—	X	—	X	—	X	—	X	—	X	—	O	X	X	—	X	—	X	—	—	—	—	—
Ex. 3	X	—	X	—	X	—	X	—	X	X	X	—	X	—	X	—	X	—	X	—	X	—	X	—	X	—	X	—	—	—	—	—
Ex. 4	X	—	X	—	X	—	X	—	X	—	X	—	X	—	X	—	X	—	X	—	X	—	X	—	X	—	X	—	—	—	—	—
Ex. 5	X	O	X	O	O	O	O	O	X	O	X	O	O	O	O	O	X	O	X	O	O	O	O	O	X	O	X	—	—	—	—	—
t; t-32	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	34	25	26	27	28	29	30	31	32
Legend						X Struck Note								— Held Note									O Optional: Struck or Held									

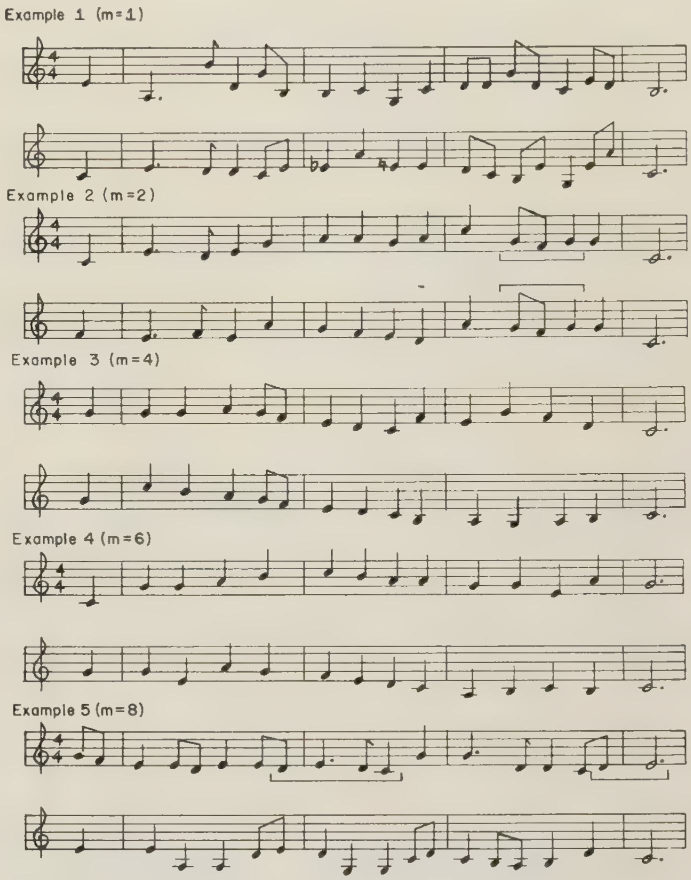
while the $(2m-2)$-gram and the $(2m-1)$-gram extend across at most m quarter notes, the $(2m)$-gram and the $(2m+1)$-gram each extend across at most $(m+1)$ quarter notes. Examples of this phenomenon are shown in Table IV, while the counts of m-grams in Table II, having similar extension properties, are joined with dashes.

The synthesis for $m=1$ is random, obeying the (monogram) probability distribution. Any generated held note was interpreted as a continuation of the preceding struck note. A monogram hymn is given in Example 1 (right). The particular constraint used for this hymn is shown in Table V, along with constraints for the following examples. The low probability of accidentals (sharps and flats) in the sample, less than one per cent, resulted in the presence of only one accidental in Example 1: the *E* flat for the forty-third and forty-fourth notes. Despite this, however, the hymn is not easy to sing and contains unnatural intervals.

The digram hymn given as Example 2 exhibits several interesting irregularities. The constraint is a dotted constraint with two optional eighths (see Table V). Although the trigram *G F G* exists nowhere in the sample, this combination appears twice in the example as indicated by the brackets. In both cases, the optional cell contains the struck note *F* even though the held note $\overline{G}$ is much more likely to follow the *G*. Finally, nowhere in the sample does a *G* precede a *C* dotted half at the end of a phrase, nor does a second phrase begin on *F* after the first phrase ends on *C*. Indeed, none of these features are in keeping with the usual explicit rules of composition, but they are permitted by the inadequacy of the low order of the analysis-synthesis.

In the syntheses of orders 4 and 5, there is less roughness of the generated hymn. A tetragram hymn with no options in the metric constraint (Table V) is shown as Example 3. The problem of excessive range is one which was introduced implicitly by the naive method of transposition. Most of the original hymns had melodic lines with a range of about an octave, but the transpositions spread these ranges away from the normal vocal range.

The tetragram hymn illustrates a subtle manifestation of the first synthesis hindering effect, one that was not anticipated. In syntheses of intermediate order, there were long ascending or descending sequences each made up of a succession of the short ascents or descents so common in the sample. With higher-order procedures, these overlong sequences cannot occur.



In Table III it is seen that the yield with $m=7$ represents a minimum for each constraint and that the yield with $m=6$ is quite near this minimum. A hexagram hymn generated with a basic quarter-note constraint is given in Example 4. This hymn demonstrates the existence of the "middle ground," and nowhere contains more than four consecutive quarter notes of any hymn in the sample. It shows the long-descent effect to some degree.

The yield for $m=8$ in Table III is appreciably greater than the yield for $m=7$. An examination of the octogram hymns reveals that a few of them are wholly identical with hymns in the sample. Several others have the entire first phrase of one hymn in the sample and the entire second phrase of another. An output hymn of order eight, which is an interesting composite of three hymns, is given as Example 5. The constraint is of the skeletal type and is shown in Table V. The seven-note section in brackets is common to two hymns at the seg-

ments and permits the changing of hymns in midphrase in this otherwise degenerate case. In a large number of output hymns in which such segmentation occurs, the transition preserves the absolute time coordinate of the original hymn. This was not true for orders lower than the eighth. Hence, at the eighth order, the third synthesis-hindering effect originally predicted has appeared. The sample is so redundant that the synthesis will not yield "original" utterances when carried out beyond the eighth order.

The skeletal constraint managed to produce a greater yield than the quarter note constraint for $m=8$. On the lower orders of synthesis ($m=4$ and 5), the former constraint permitted the hymn to run so astray that it could not meet a subsequent constraint. It hence produced fewer acceptable hymns than the quarter-note constraint. For $m=8$, however, the implicit structures revealed by the octogram analysis were strong enough to keep the synthesis within the framework of the sample, thus producing mostly hymns of which each half was exactly like half of a sample hymn. This is a further result of degeneracy.

Extension

The present experiments have permitted the identification and characterization of the limiting effects that apply to any generalized analysis and synthesis, any induction and subsequent deduction. Since these processes will become more and more important in the application of computing machinery to more delicate and sophisticated tasks, it is desirable to explore their characteristics and properties more fully. It is important to develop some measure of structural complexity in terms of information content, to develop better ways of characterizing the extent of limitation imposed by constraints, and to develop methods of describing the information content of a sample which is sufficient to permit synthesis of original members of some structure class.

Ideally, the present experiments would not have been needed. One would have described the information content of the sample and the extent of the constraints of the musical structure, and from this one could have predicted that first and second-order Markoff analysis-synthesis would have yielded sequences unacceptable as hymns and that eighth order analysis-synthesis would have yielded sample members.

The wide discrepancy between the ideal situation and that which currently prevails emphasizes the large amount of theoretical and experimental work which will have to be done before the inductive-deductive processes are well enough understood for general use in computing machine applications.

Acknowledgment

The authors are indebted to their colleagues at the Harvard Computation Laboratory for advice and encouragement. Profs. K. E. Iverson, R. C. Minnick, and A. G. Oettinger gave especially valuable assistance. We are also indebted to Prof. H. H. Aiken for making available the Laboratory facilities for this investigation.

An Electronic Analog Cross Correlator for Dip Logs*

J. H. SASSEEN†

Summary—**The Dip-Log Correlator is a special purpose electronic analog computer which rapidly cross correlates the three variable functions of a dip-meter well log. This cross correlator utilizes a novel derivation and application of the cross-correlation function from information theory. A prerequisite of the system is the recording of the dip-meter data on magnetic tape for computation by the cross correlator. The triple cross-correlation technique and some of the essential circuitry of the cross correlator which is in operation as a geophysical tool at the Humble Oil & Refining Company Houston Research Center are described.**

Introduction

In recent years, some of the most important advances in the broad field of communications are the outgrowth of information theory studies. Information theory is a fundamental mathematical study of the nature of information. It is considerably involved with statistical theory. Among other things, information theory has been applied to radio communication, radar, acoustics, meteorology, and even economics.

Among the most useful tools from information theory are the correlation functions. These functions are applicable to the data obtained from dip-meter well surveys since the computations involving these data call for recognizing the similarity of events between the three dip-meter functions and determining their "best fit" by relative shifts with respect to each other. Because these data are collected and presented in an analog form, an electronic analog computer seems logical for performing these correlation operations.

Dip-meter well logging or surveying involves measurements from a bore hole in the earth such that the dip

* Manuscript received by the PGEC, April 28, 1957.
† Humble Oil & Refining Co., Houston, Texas.

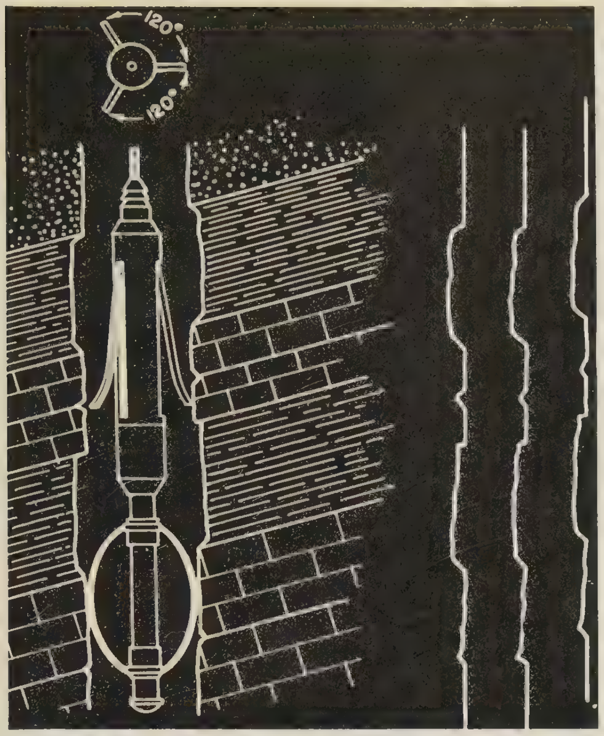


Fig. 1.

magnitude and direction (with respect to true north) of a bedding plane or formation in the earth may be calculated. The measurements involved may be of any physical property which varies from formation to formation such as resistivity, self-potential, or bore-hole radius. The selected property is measured independently along three separate vertical paths. These paths are spaced 120° apart and extend the entire length of the bore hole (Fig. 1). As the dip-meter device is drawn up the bore hole by means of an electrical cable, three arms carrying the measuring devices are pressed against the walls of the bore hole enabling the formation characteristics to be recorded at the surface. Thus, three curves are produced as illustrated. Results depend to a large extent upon a break or change occurring at the formation boundaries. The apparent dip of a bed or formation may be calculated by measuring the displacement between similar events or breaks on these three curves in terms of bore-hole distance. In order to obtain the true dip magnitude and direction, the hole diameter, the amount and direction of hole inclination, and the direction of one of the measuring arms with respect to north must be known. This information is also continuously recorded by the dip meter.

The difficult problem in dip computation is the correlating or picking of these breaks in order to determine the relative displacements between the measured dip-log data. These data vary in quality from good or highly coherent information, as shown, to poor or slightly coherent. In general, the data are not as coherent as those which are shown and the human computer has trouble in correlating the data. It is also evident that the correlation of the data by a human computer is a highly subjective process. For this reason, results are likely to vary widely from one computer to another, particularly when the data are poor. An electronic analog cross correlator has been designed and constructed to aid the human computer in correlating the three curves of dip-meter well surveys.

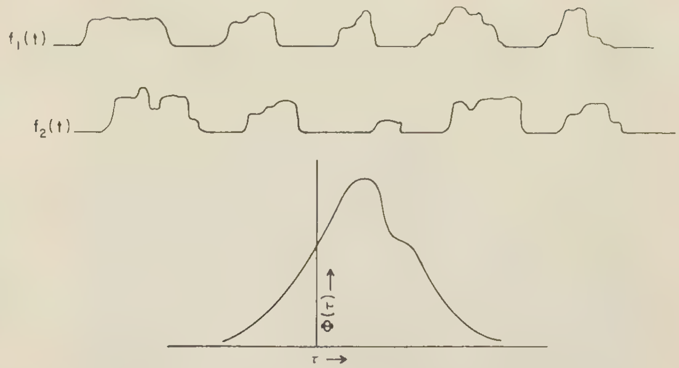


Fig. 2—Cross correlation:

$$\Phi(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{+T} [f_1(t)][f_2(t+\tau)]dt.$$

Cross-Correlation Functions

The dip-log cross correlator operates on the three dip-log curves and performs the mathematical equivalent of a triple cross-correlation function. Let us consider the method of cross correlating only two functions of a common independent variable (Fig. 2). Illustrated are two dip-log data functions of the common independent variable, bore-hole depth, an equation for the cross correlation of two functions, and the cross-correlation curve which shows the degree of statistical similarity existing between the two functions. In essence, the equation states: take the average of the product of the two functions, for the interval $-T$ to $+T$, with the two functions displaced or shifted by an amount τ. (Actually, the equation calls for the average over the interval minus infinity to plus infinity. In practice, of course, an approximation is necessary.) This procedure is carried out for all τ's or displacements along the independent variable axis, and the resulting correlation curve is plotted. The maximum on this curve represents the point of optimum correlation. The dip-log cross correlator obtains and combines the cross-correlation curves of the three possible pairs of the three dip-log data functions; that is,

$$\phi_{1,2,3}(\sigma, \tau) = \phi_{1\,2}(\sigma) + \phi_{1\,3}(\tau) + \phi_{2\,3}(\tau - \sigma)$$

where σ is the distance that data function $f_2(S)$ is shifted with respect to $f_1(S)$ and τ is the distance that data function $f_3(S)$ is shifted with respect to $f_1(S)$. In the above equation,

$$\phi_{1,2}(\sigma) = \overline{f_1(S)f_2(S+\sigma)}$$

$$\phi_{1,3}(\tau) = \overline{f_1(S)f_3(S+\tau)}$$

$$\phi_{2,3}(\tau - \sigma) = \overline{f_2(S+\sigma)f_3(S+\tau)}.$$

The dip-log cross correlator obtains the sum of the cross products and thus the triple cross-correlation function $\phi_{1,2,3}(\sigma, \tau)$ in a rather simple way. Consider the expression

$$\overline{[f_1(S) + f_2(S+\sigma) + f_3(S+\tau)]^2}.$$

This expression when expanded results in a sum of terms as follows:

$$\overline{2f_1(S)f_2(S+\sigma)} + \overline{2f_1(S)f_3(S+\tau)} + \overline{2f_2(S)f_3(S+\tau-\sigma)} + \overline{f_1(S)^2} + \overline{f_2(S+\sigma)^2} + \overline{f_3(S+\tau)^2}.$$

It may be seen that the sum of the first three terms results in an expression which is proportional to the triple cross-correlation function $\phi_{1,2,3}(\sigma, \tau)$; in addition, there are three squared terms which are undesired. Each of these squared terms represents the mean power contained in the data functions f_1, f_2, and f_3, respectively. If a sufficiently long statistical sample is taken, then these squared terms will not be functions of σ or τ, and consequently, the sum of the three terms may be considered constant. Thus the variable portion of the above expression is proportional to the desired triple cross-correlation function $\phi_{1,2,3}(\sigma, \tau)$.

It is evident that the electronic analog section of the correlator need only shift two of the data functions independently with respect to the third and add the three data functions, square this sum, and then obtain the average over a suitable interval. In the correlation process of the correlator, the squared terms are held equal and constant, for all practical purposes, and are taken into account in calculating the correlation index. The dip-log cross correlator does not plot the surface or family of curves generated by the triple cross-correlation function $\phi_{1,2,3}(\sigma, \tau)$, since the entire surface is not required and to do so would slow up the operation. Only the value of σ and τ at the maximum or peak correlation is required, and a meter is used to indicate this maximum.

Dip-Log Cross-Correlator Equipment

A prerequisite of the correlator system is the recording of the dip-meter data on magnetic tape for computation by the cross correlator. The magnetic-tape equipment for recording dip-meter signals in the field consists of a set of recorder electronics, its power supply, and a 14-track magnetic-tape transport (Fig. 3). This recording system connects into the output circuits of any of the conventional dip-meter loggers. The recorder electronics is a direct recording system which essentially consists of a 3.5-kc bias oscillator and three push-pull dc amplifiers for the dip-meter signals. This 3.5-kc bias and the dip-meter signals are added together in the 50-mh center-tapped magnetic record heads. Also included in the recorder unit is a pulse system for recording depth-mark pulses and pulse-type orientation data (hole inclination, its direction, etc.). The magnetic-tape transport utilizes tape that is one inch wide. This magnetic tape is driven

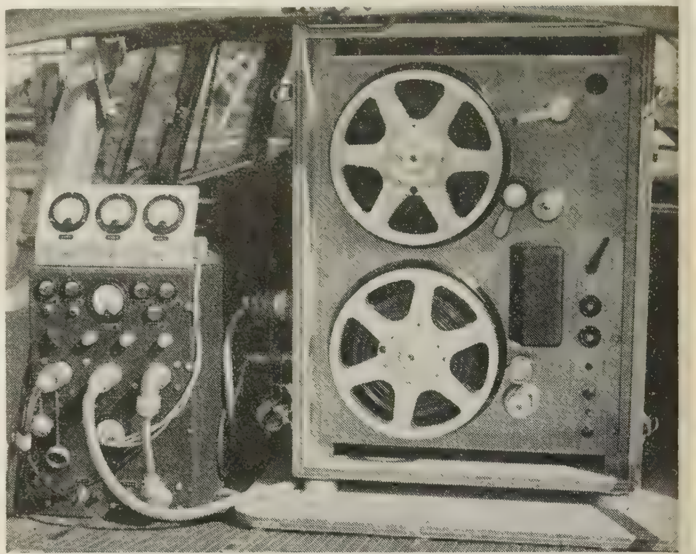


Fig. 3.

by a selsyn system which follows the movement of the logging cable. A one-foot movement of the logging cable results in six-hundredths of an inch of magnetic-tape movement; thus, at a logging rate of 60 feet per minute, the tape speed is six-hundredths of an inch per second. Hence, only 50 feet of tape is required to log a 10,000-foot well. This contrasts with the 1000 feet of paper or film for the usual visual record. The dip-meter magnetic-tape recording is brought in from the field to the dip-log cross correlator to be computed.

The dip-log cross correlator is composed of a number of units. Fig. 4 is a simplified block diagram of the equipment. Essentially, two operations are performed: first, the three channels of dip-log data are simultaneously transcribed section by section from the field magnetic-tape recording onto the magnetic storage drum; next, these data are correlated by means of the electronic analog computer. The playback of the signals for transcription is performed at constant tape speed; therefore, time becomes the analog of depth in the system. This playback speed is about 1000 times faster than the speed at which the signals were recorded, and hence all recorded frequencies are projected upwards a thousandfold. In other words, at a logging rate of about 60 feet per minute, a 45-cps signal becomes 45 kc, the upper frequency limit of the system, and a 0.25-cps signal becomes 250 cps, the lower frequency limit of the system. The transcription process for the signals from the tape to the magnetic storage drum is also a direct-recording system. In order to transcribe data to the drum, the correlator operator sets a number into the preset counter corresponding to the starting depth of the data to be correlated. Then, a cycling button is pressed, which automatically initiates a series of events as follows. First, any previous signals on the storage drum are erased. Second, the tape-transport mechanism starts and comes up to the speed of 60 inches per second. Third, the counter counts depth pulses until the preset count is reached, at which time the transcription-

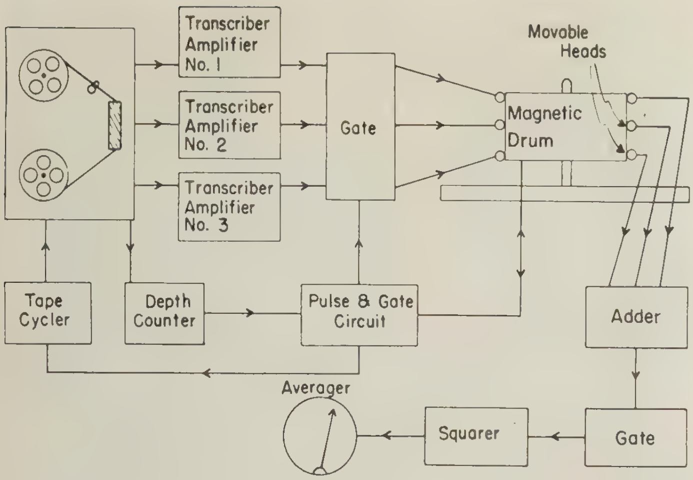


Fig. 4—Office transcription and correlation equipment.

amplifier gates are opened, allowing data to flow from the tape to the drum. After one complete revolution of the drum, the gates are closed and the drum contains 100 feet of dip-meter data. Actually, about 113 feet are recorded to allow leeway in gating. Fourth, after the section of data has been recorded, the tape transport automatically reverses and rewinds to the beginning of the record where it stops, ready for the next transcription cycle. The tape is stopped after rewind by means of a photocell circuit, which is actuated by an appropriately placed hole in the magnetic tape. This hole is previously punched ahead of the actual data on the tape to allow the tape to accelerate to full speed before the data are encountered. This transcription cycle requires about eight seconds on the average. However, it is not necessary for the operator to wait this entire time before starting to correlate the data, since part of the eight seconds is consumed in rewinding the tape after the data have been transcribed. A complete transcription circuit consists of a preamplifier, an equalization network, a constant-current output stage, and a bias oscillator with its power-amplifier stage.

With signals on the high-speed magnetic drum, the major electronic steps in the cross-correlation process are then as follows. First, the electrical signals from the three dip-log data channels on the drum pass to three preamplifiers and thence to an adding circuit. From this point on, only one channel is required. Second, the resulting signal is then equalized to compensate for the frequency response characteristics of the playback head and drum. Third, the signal next passes through the gating relay which is cycled by the pulse and gate circuit in accordance with the setting of the gating heads which are mounted around the drum. Fourth, the signals pass to a power amplifier and then to a squaring and averaging circuit. This squaring and averaging operation is carried out by means of vacuum-type thermocouples. Fifth, the output of these thermocouples is then amplified by means of a chopper stabilized dc amplifier and indicated by a meter.

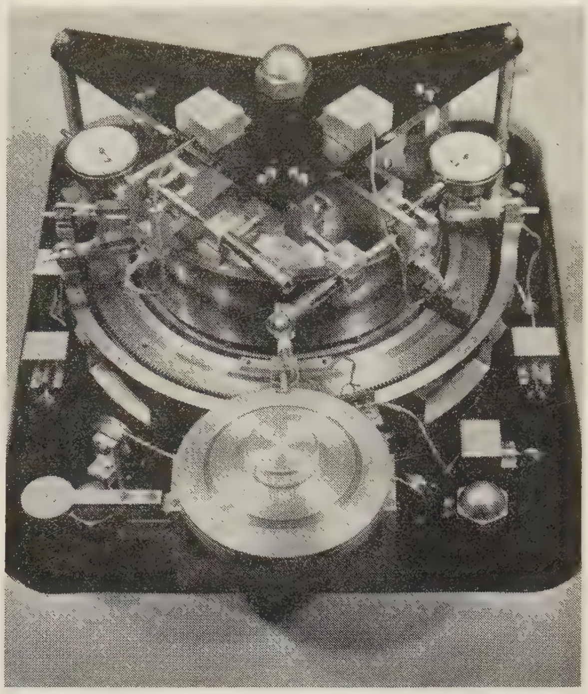


Fig. 5.

One of the major units of the correlator is the magnetic drum which provides a storage medium for the computer (Fig. 5). This ten-inch-diameter drum is turned about 530 rpm or 280 ips and utilizes magnetic heads which are spaced away from the drum by approximately 0.0008 inch. Nonferrous materials were used as much as possible in the construction of the drum housing. The drum is made of an aluminum alloy forging and it is mounted on precision bearings which have very small eccentricities. The final cut on the drum surface was made with the drum turning on these bearings. The outage of the drum is less than ± 50 microinches. The three-inch-wide drum was then very uniformly sprayed with many coats of magnetic oxide, the final coating thickness being less than one-thousandth of an inch. Four types of magnetic heads are used on the drum—dip-log data record heads, dip-log data playback heads, pulse record heads, and pulse playback heads. The erasure of each track is accomplished by using the record heads as erase heads. The pulse channels are erased with dc, while the dip-log data channels are erased with a 150-kc signal. Considerable ac power is required for erasure at 150 kc because of the core losses in the magnetic heads. This ac power is permitted to flow to the heads for only about 0.3 second—just long enough to allow the drum to rotate about three times. Thus, the heating of the heads is kept to a minimum.

Two of the three dip-log playback heads are movable. Each of these heads is attached to a micrometer barrel for movement around the periphery of the drum. The heads can be moved plus or minus one-half an inch, which corresponds to a plus or minus 20-inch vertical displacement of the dip-log data curves in the bore hole. A calibrated indicator also attached to the micrometer

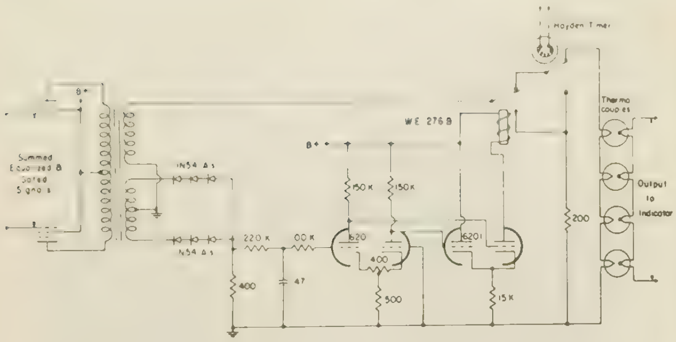


Fig. 6—Thermocouple protective circuit.

provides a direct reading of the equivalent vertical displacement of the dip-logging measuring device in increments of four-hundredths of an inch. Two of the pulse playback heads are movable through about 330° of the drum periphery, and their setting determines the gated fraction of the 100 feet of data. These heads may be moved independently, or they may be moved as a pair around the drum. This is accomplished by means of a clutch and gear box that drive the ring gears to which these pulse heads are attached. Scales are provided for reading the number of feet in the interval being gated for correlation and for location of this interval in the 100 feet of data on the drum.

One of the most interesting electronic circuits of the cross correlator is the squaring and averaging circuit (Fig. 6). The squaring and averaging is accomplished by means of vacuum-type thermocouples which are protected from "burn-out" by the associated protective circuit. Some power is required to drive the thermocouples, and this is furnished by a power-amplifier stage that is capable of about 10 to 12-w peak power. The rated average power for the four thermocouples is about 16 milliwatts. The 10 to 12-w peak power capability of the power amplifier is necessary because of the extremely high peak-to-average power ratio of the gated dip-meter signals. This ratio can run as high as 3000 to 1 under extreme conditions. Thermocouples square and average quite well in spite of the high peak-to-average power demand, since their action is independent of peak values. Also, the thermocouples' squaring action is almost perfect. The thermocouples are operated at an average power level of about 1 milliwatt or less in order to allow for high peak-to-average power ratios and to provide a "burn-out" safety factor. Since the thermocouples will only tolerate a relatively small overload, they have been provided with a protective circuit. This circuit uses a crystal-diode power monitor which continuously monitors the signals to the thermocouples. The output of this monitor is fed to a direct-coupled differential amplifier and then to the relay tube. This relay acts to open the circuit feeding the thermocouples whenever the average power exceeds a preset value. The time constants of the protection circuit are such that the thermocouples are protected from either slow or suddenly applied overload signals. Although the crystal-diode monitor circuit is not perfectly square law, its action is adequate in the protective circuit. A time-delay relay is used to protect the thermocouples from possible overload during the warm-up period before the protective-circuit heaters have had a chance to reach equilibrium temperature.

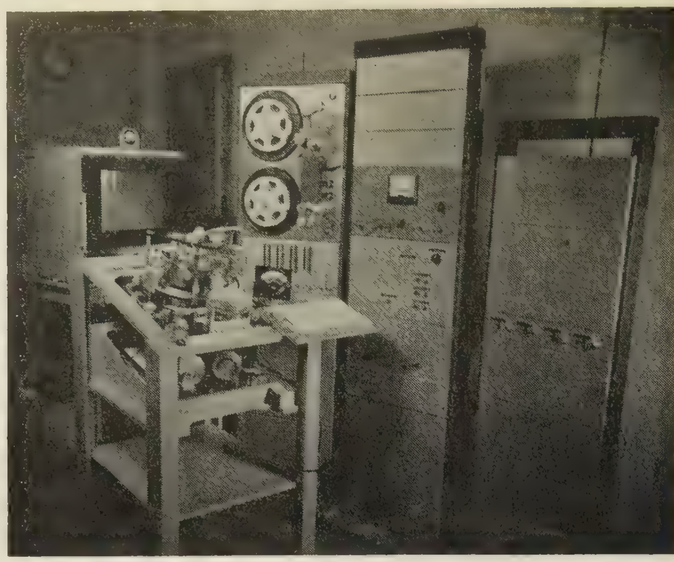


Fig. 7.

Although some of the circuits of the dip cross correlator are rather complex, the operation of the equipment is simple. Fig. 7 shows the cross-correlator equipment set up for operation. The high-speed magnetic drum, its accessories, and the correlation meter are mounted in the table-like framework. The three relay racks contain the playback tape transport, power supplies, circuits for transcribing data to the drum, gating circuits, and the necessary adding, squaring, and averaging circuits. A five-trace 21-inch oscilloscope is used for viewing the three dip-meter profile traces, along with a sum and correlation interval trace and a bore-hole depth trace.

In operation, any 100-foot section of the log may be transcribed to the correlator drum and any interval from 8 feet to 100 feet may be correlated. Correlation is performed by alternately turning the micrometers on the two movable magnetic heads. As this is done, the output of the analog computer is observed on the correlation meter for a maximum reading. At this meter reading, the dip displacements are read from the calibrated micrometer dials. Thus, the displacements needed to compute the dip are obtained at a "best-fit" point at the maximum of the triple cross-correlation function. Also, from the meter reading, the per cent correlation or correlation index for the interval is easily computed. This is a measure of the degree of similarity between the three dip-log data functions. A 100 per cent correlation index is obtained only if the three profiles are exactly identical. This first correlator model has not been made fully automatic. Therefore, it is necessary to set the dip displacements along with the orientation data into a mechanical analog dip resolver to obtain the true dip.

Conclusion

The dip cross correlator possesses several advantages over manual or human computation. Actually, some indirect advantages are realized by merely recording the data on magnetic tape whereby small, weak signals can be amplified to a usable level on playback. Also, a magnetic-tape recording exhibits a greater dynamic range and a wider frequency response than the conventional film and paper recordings. Direct advantages of the correlator are as follows. First, the correlator which is completely objective computes a perfectly repeatable answer obtained by the "best-fit" criterion of the triple cross-correlation function. This statement does not imply that one answer is always obtained, for the correlator can give two equally weighted answers for an interval. However, this is a rare occurrence in comparison to the frequent multiple answers that can be obtained by manual computation. Second, with each answer, the correlator produces mathematical grades that are a measure of the similarity between the data traces. Third, the correlator can compute signals in the presence of incoherent noise; in fact, this type of noise can be as large or larger than the signal. Fourth, the correlator computes data regarded as "impossible" or "no good" by the human computer. Fifth, the correlator for two dip-logs from the same well produces a better agreement between the two sets of results than the human computer. Sixth, at present, the correlator with good data is slightly faster than the human computer. However, it is the exceptional log that has these good data. The human computer must spend his time in inverse proportion to the quality of the data. Thus, the time relation between the correlator and the human computer is a variable; with poor data, the correlator can be as much as ten times faster than the human computer. Perhaps a utility, but not another advantage, of the correlator is the fact that it is an extremely useful research tool in the analysis of data, autocorrelations, power-density spectra, etc.

The evaluation of the results of the dip-log cross correlator is difficult, since no true performance criterion exists. That is, the dip of the formations that surround a bore hole in the earth is not always known. However, in some instances, the geological and geophysical control is good enough to allow a direct comparison with the correlator results. In most cases, though, the comparison must be made between the correlator results and those of the human computer. The results obtained to date are encouraging and the correlator has proved to be a useful geophysical tool. In some instances, the dip correlator has computed reliable results from data which could not be worked by a human computer.

A Variable Function Delay for Analog Computers*

R. S. STONE† AND R. A. DANDL†

Summary—**Analysis of reactor problems requires the simulation of transport lags. This paper discusses the design of a compact and relatively simple transport lag utilizing a synchronously-switched bank of condensers.**

THE major controls problem at the Oak Ridge National Laboratory; *i.e.*, the safe operation of a nuclear reactor, is in general well suited to routine analysis in the analog form. There are, however, certain reactor characteristics which are troublesome to simulate. One commonly encountered difficulty involves the transport of a spatially-defined parameter—temperature, for example—from one portion of the system to another, as in flow through a pipe. Finite velocity laminar flows require multisecond high-fidelity delays, which are difficult to arrange in electrical systems.

Until recently, the analog circuit used for transport lags at this laboratory was an RC time delay based upon Taylor's expansion of the function $f(t-T)$. While adequate for delaying transients with periods at least ten times the delay time, this system introduces successively greater amplitude and phase distortion as the period of the delayed transients becomes comparable with the delay time.

Other workers have used either a faster converging series or direct network analysis to develop a more effective approximation to the delay function.[1–4] The resultant network, in which a lumped constant LC delay line of some complexity is simulated,[5] has practical limitations because of the large number of amplifiers in-

* Manuscript received by the PGEC, May 3, 1956; revised manuscript received, July 11, 1957.
† Oak Ridge Natl. Lab., Oak Ridge, Tenn.

[1] C. D. Morrill, "A sub-audio time delay circuit," IRE Trans., vol. EC-3, pp. 45–49; June, 1954.
[2] W. J. Cunningham, "Time delay networks for an analog computer," IRE Trans., vol. EC-3, pp. 16–18; December, 1954.
[3] W. E. Thomson, "Time delay circuits," IRE Trans., vol. EC-4, p. 74; June, 1955.
[4] C. H. Single, "An analog for process lags," *Control Eng.*, vol. 3, pp. 113–115; October, 1956.
[5] A. J. Ferguson, "A note on phase correction in electrical delay networks," *Can. J. Res.*, vol. 25, sec. A, pp. 68–71; January, 1947.

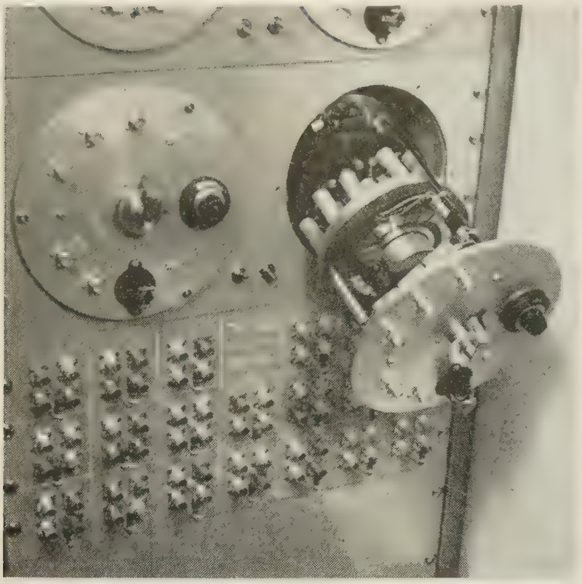


Fig. 1—Switch assembly and mounting.

volved. A circuit of this type requires at least three operational amplifiers when the delayed function has component frequencies with periods comparable in length to the delay. With more stringent response requirements, the number of amplifiers goes up proportionally. A few lags of this type will quickly use up the amplifier resources of a small facility.

Magnetic tape delays,[6] though generally lacking in dynamic range, offer wide bandwidth and a broad and easily changed range of delay times. The desirable operational features are unfortunately compromised by the dollar and space requirements imposed by a respectable number of such machines.

Since problems under consideration indicated the need for about six time delays, the development of a compact, relatively simple transport lag seemed worth while.

The device shown in Fig. 1 is a synchronously-switched bank of condensers which store a sampled version of the input function. A writing brush contacts each condenser in sequence and impresses on it a voltage corresponding to a point on the time-varying function which one wishes to delay. Fig. 2 is a schematic of a 20-contact transport lag. The amplifier switch configuration shown performs a 19-point approximation of $-\frac{1}{2}f(t-T)$ when operating on $f(t)$. In the case of an N contact switch, T equals n/NF seconds, where F is the rps of the brushes and n is the number of contacts by which the read brush lags the write brush. ΔC is a relatively small capacitor to minimize switching transients.

Operation of the circuit may be described through reference to the functional representation in Fig. 3. In the writing case one wishes to obtain a transient solution for the voltage e_w to which C_w (one of N identical

[6] D. C. Reukauf, "Simulate transport lags with magnetic tape," *Control Eng.*, vol. 4, pp. 145–147; June, 1957.

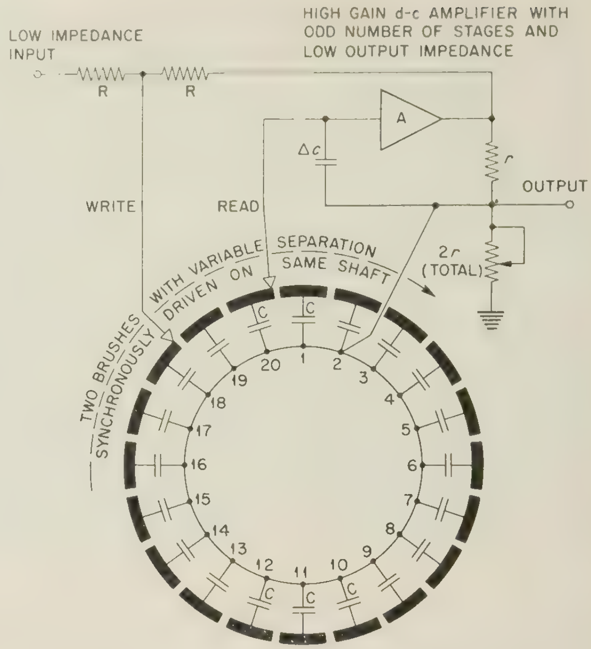


Fig. 2—Wiring diagram of function delay switch.

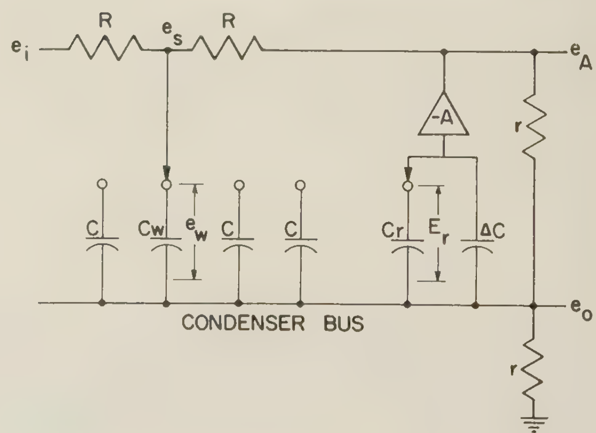


Fig. 3—Operational diagram of function delay.

capacitors) is charged to be read T seconds later. The nodal equations are as follows:

$$\frac{e_i - e_s}{R} - \frac{e_s - e_A}{R} = PC(e_s - e_0) \tag{1}$$

$$\frac{e_A - e_0}{r} + PC(e_s - e_0) = \frac{e_0}{r} \cdot \tag{2}$$

Subtracting (1) from (2), and solving for $(e_r - e_0) = e_w$, gives

$$e_w = \frac{e_i}{P(RC + rC) + 2} \cdot \tag{3}$$

When $e_i(t)$ is a step of magnitude E_i, (3) may be operated on to yield

$$e_w = \frac{E_i}{2}\left(1 - \exp^{-2t/(R+r)C}\right). \tag{4}$$

This shows that for

$$t \gg \left(\frac{R+r}{2}\right)C, \qquad e_w \cong E_i/2$$

and is independent of e_0, the voltage being read.

The amplifier output is related to the constant voltage on the capacitor being read, by

$$e_A = -A(E_r + e_0). \tag{5}$$

This shows that as A approaches ∞, e_0 approximates $-E_r$ for the time ΔT during which C_r is being read so that the read voltage is independent of that being simultaneously written "upstream." Since E_r, the voltage across C_r, was the e_w of n contacts previously,

$$E_r = E_i/2 \text{ delayed, or} \tag{6}$$

$$e_0(t) = -1/2 e_i(t - T). \tag{7}$$

The effect of circuit parameters upon the behavior of the system is as follows. The plastic impregnated paper capacitors currently used in the switches have time constants of the order of 10^5 seconds. This means that a 100-second delay between read and write produces a maximum leakage error of 0.1 per cent.

Amplifier grid current also causes loss of charge from the condensers. It can be shown that to limit such voltage drops to 1 millivolt or less, with grid current of 10^{-10} amperes,

$$C \geq 10^{-7}\Delta T, \text{ where}$$

$$\Delta T = 1/NF = \text{dwell time/contact}. \tag{8}$$

From (4), it is apparent that the relative error due to hasty writing is $\exp -2\Delta T/(R+r)C$. For a maximum error of 0.1 per cent,

$$\frac{(R+r)C}{2} \leq \frac{\Delta T}{6.9}. \tag{9}$$

Solution of (8) and (9) for T/C yields

$$\frac{6.9(R+r)}{2} \leq \frac{\Delta T}{C} \leq 10^7. \tag{10}$$

If maximum brush time per contact is two seconds, the right-hand portion of (10) indicates the adequacy of 0.22-microfarad capacitors. If minimum ΔT is 0.3 second and $C = 0.22$ microfarad, the left-hand portion of (10) requires that

$$R + r \leq \frac{2 \times 0.3}{6.9 \times 0.22 \times 10^{-6}} = 395{,}000 \text{ ohms}. \tag{11}$$

A further restriction on R and r is the requirement that they be of such values as will prevent transient overloading of the operational amplifier. The values necessary are a function of the output stage involved, but typical values for a cathode follower limited to 20-milliamperes peak current and working between $+300$ v and -300 v, are $r = 10{,}000$ ohms and $R = 100{,}000$ ohms.

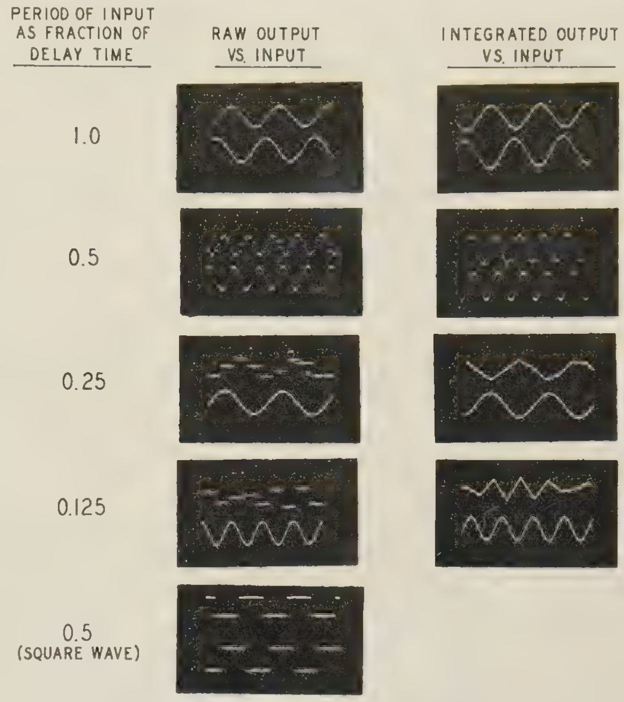


Fig. 4—Waveforms of delayed outputs for various inputs to synchronous switch. In each double oscillogram, the lower curve shows the input to the switch; the upper curve shows the delayed output.

To insure that the write brush never enters a contact before the read brush has left it, the 20-contact switch described above is run with a maximum lag of 18 contacts. At such a setting, the maximum error in delay time is $1/18 = 5.6$ per cent. The response of this 18-point delay to various frequencies is shown in Fig. 4, and is seen to provide reasonable fidelity for signals with periods as short as half the delay time. Increasing the number of contacts, of course, would improve the frequency response proportionally.

Since contact synchronization is unimportant in this system (as contrasted to schemes which switch both sides of the capacitors), it should be feasible to construct switches with a very large number of contacts whose capacitors are intrinsic with the contacts, and thereby achieve a higher fidelity transport lag. The present switches have been routinely used as transport lags in simulated reactor systems for two years and have proved satisfactory in that application. One great advantage is the ease with which the lag may be varied by simply changing brush spacing and/or motor speed. As more switches are added to the facility, it should be possible to simulate very complex transport systems with a modest number of operational amplifiers.

Acknowledgment

Thanks are due to H. J. Stripling and T. F. Sliski for assistance with the mechanical design of this device.

HUMAN BEINGS AS COMPUTERS

The four papers given here were presented April 3, 1957, in Philadelphia, Pa., at a meeting sponsored jointly by the Philadelphia Section and by the PGEC. Prof. Miller's paper was read by Dr. McCulloch—*The Editor*.

Biological Computers*

WARREN S. MCCULLOCH†

THE human central nervous system is composed of some 10^{10} neurons. Each has branches, a body, and a long, thin, output lead—called an axon. The burning of sugar supplies the energy to keep the neuron's outside 0.07 volt positive to its inside. The capacity is fixed at, say, 1 μf per square centimeter, but its resistance through the surface is highly nonlinear—very great so long as the outside is above 0.02 volt to 0.03 volt positive, but below that value, nearly a dead short circuit. Hence we see that if the voltage is reduced at any small area, current runs in freely-discharging adjacent areas, which then short-circuit, draw current, and discharge the next areas. Such is the propagated nervous impulse. It proceeds as a smoke ring of currents along the axon at a velocity determined by distributed capacity, distributed resistance, *and* distributed source of power. The velocities range from about 150 msec to 1 msec. The impulse has a rise time of less than 0.5 msec and a slightly slower decay. We speak of this impulse as all-or-none, not in the sense of its being always and everywhere of the same size, but of its size at any place being determined only by the condition at that place—not in any way depending on the strength of the disturbance that initiated it. In short, a neuron is a distributed repeater. So much for the components.

We call their effective connections "synapses." Sometimes we can identify them as places where terminal branches of axons of one neuron lie against branches or cell body of another, or end in buttons applied to them. By "effective" connections I mean only that impulses in an axon evoke or prevent impulses in the neuron on which or near which that axon ends. In a few places where we know the details of anatomy and have the proper electrical records, the best guess is that if a fine axon courses along a cell body to end on or among its branches, it will, by leaving a source of current on the body and a sink among the branches, raise the voltage of the body and so inhibit its firing; whereas a large fiber coming in from among the branches to end on the cell body will almost certainly carry impulses to it to excite it or to facilitate its excitation. But that is certainly not the whole story.

There is a second way of gating impulses by impulses. Consider an impulse in an axon of uniform diameter. Its factor of safety for propagation is about 10 to 1. Now let the axon branch into two or more smaller axons. Their surfaces per unit length may exceed that of the parent axon, but their internal resistances go up as the squares of their radii go down. Hence the branches demand more current and a longer time constant. This demand decreases the safety factor at each successive branch point. These fibers are not insulated from one another by dielectrics, but their disturbances are short-circuited by embedding them in a conductor or common ground—say salt water. This does not prevent neighboring fibers from affecting one another. Seen from this salt water, every impulse is a sink of current preceded and followed by a source. Sources in neighboring fibers raise the threshold in a given fiber, and sinks lower it. In this manner, impulses in neighboring fibers help or hinder the passage of impulses to or toward the synapse with the next cell. In fact, we have ample proof that this does happen in ordinary nervous activity. An impulse dying in a fiber certainly tends to block the next farther back.

What we have said so far insures that those neurons in a brain can be and are used as relays in a computer to gate all-or-none impulses; but it says nothing of coding. We should not attempt that question until we have a better picture of the function and general organization of the central nervous system. Its function is legion, but of one general kind. Every cell in the human body has a multiple inverse feedback—giving it great dynamic stability. Most of our juices are buffered. This is necessary if our parts are to survive in a changing environment. But, our body as a whole would die promptly were it not for our reflexive mechanisms, each an inverse feedback. Of these there are certainly more than a thousand but less than a million distinguishable loops. Many pairs of such loops acting simultaneously would produce contrary effects (like the motion of the knee in opposite di-

* Manuscript received by the PGEC, May 14, 1957. This work was supported in part by the U. S. Army (Signal Corps), the U. S. Air Force (Office of Sci. Res. and Dev. Command), and the U. S. Navy (Office of Naval Res.).

† Res. Lab. of Electronics, M.I.T., Cambridge, Mass.

rections) or disastrous results (like inhaling and swallowing). Always, many have to be employed in any complex act. Hence the principal function of the central nervous system is that of computer for an enormously complex, multiple closed-loop servosystem. All this is too general to be helpful. Let us therefore look at a bump at the back of the hindbrain, called the cerebellum.

The chief job of the cerebellum is to bring to rest at the right place whatever part of the body is put in motion. To do this it has first to be informed of the position, velocity, and acceleration of each part, which is done by paths ascending from the parts, relayed in the spinal cord. Next it needs to know the program and subroutines involved—information that it gets from the basal ganglia and the brain stem. In as much as the act occurs in a gravitational field and generally requires knowledge of the rotation and translation of the head, the cerebellum has to be informed by that part of the ear which handles this sort of information—the so-called labyrinth. Finally, the cerebellum needs information from distance receptors by light and sound as to the world about it and the objects in it. This information has to be processed by the bark of the big brain, the cerebral cortex, before it is sent to the cerebellum. Having calculated the necessary orders to the muscles to bring things to rest at the right places, the cerebellum sends them to the so-called reticular formation of the brain stem, which relays them to motor nuclei to be combined with local information to order the muscles to contract or relax in proper fashion.

The reticular formation plays an important role in this regulation. Consider any multiple closed-loop servosystem, and suppose that each of its component loops is stable but that all control any single effector—an aileron or a foot. Now it can easily happen that the combination jams the system at one end of its range or throws it into oscillation, which may be schizogenic. To prevent this happening it is now common practice to take part of the input derived from each loop to a common adder and feed it back inversely to many or all of the output devices. In the human body the outer and upper part of the reticular formation receives impulses from all of the ascending systems of information, and its descending fibers tend to facilitate all activity; but they also play on the central, inner part of the reticular formation, which receives the bulk of its input from descending systems coming from cerebrum, basal ganglia, and cerebellum. The output of this reticular formation is chiefly down the spinal cord to inhibit all neurons from exciting muscles. Its control is so powerful that it can stop a convulsion produced by strychnine. This is a massive, crude effect of a generalized inhibitor. As yet we have no idea how it is normally coded. We only know that it requires repetitive firing to be grossly effective.

Since the nervous system is a parallel, as well as a sequential system of computers, coding can be of two kinds. One involves which fibers are carrying the impulses; the other, the timing of impulses in fibers. The first depends largely on the anatomy, the second on the physiology of the synapses. Take them in that order. There are a couple of million axons from eye to brain, say fifty thousand from ears, and about one million from the rest of the body, all told. These preserve some semblance of their relative positions all the way from the surface of eye, ear, and body, through successive relays to the bark of forebrain, midbrain, and hindbrain. It is enough to insure that neighborhood of sense organ is converted into neighborhood in representation throughout sensory systems. Thus the topological relations of inputs are preserved (a trick that we should remember in dealing with nonnumerical data) for a ten-cent lens is about a million parallel channels, and even if it distorts or blurs the picture, it is still safer to trust its images than to trust our present attempts to scan and reconstruct, as in radar and television. In this sense we are as much true space computers as in our servosystems we are true time computers. Projection of neighborhood into neighborhood renders the system relatively stable under perturbation of threshold of cells, strength of stimuli, and detail of connection, even under scattered loss of channels—be they axons or cell bodies. These advantages make the system resemble the analog devices wherein most errors are in the last place, rather than digital devices using any radix, wherein the error is as likely to affect the first as the last digit. So much for the anatomy.

Now for temporal coding! When Pitts and I constructed the first calculus for neuronal nets, or repeater nets, we took it for granted that we could pass only one bit per impulse, that is, about one per millisecond. But increasing knowledge of fibers showed that with an axon receiving a constant stimulus, it emitted a spike with a delay that scattered only some 20 to 30 μsec, and that two impulses that summed fully at 120 μsec failed to sum at all at about 160 μsec. Hence the reliability with which it detects and emits by pulse interval modulation. This, MacKay and I calculated—showing that in this manner one could pass far more information at far lower duty cycles than we had previously expected. Alexander Andrew then showed that still more could be passed by better coding, making more use of shorter intervals. Then came our work on the input neurons. It showed that the maximum rates that could be got past its branch points was far less than had been expected. We presented that work, as a more realistic upper bound on information capacity, at the London Symposium on Information Theory. I mention this specifically because our difficulty was this: interference would cause omission of signals in neighboring channels, most when they were closest in time and represented points most closely located in muscle or skin. No one yet has been able to optimize the code for these kinds of noise, so we cannot put a reasonably high lower bound on information capacity. Amassian has shown that some cells in the cerebral cortex code into pulse interval modulation in-

formation as to the place whence their excitation came. This cortex is the organ that abstracts the ideas of objects about us from sensory inputs of any kind.

Let us, therefore, turn to some things we do know about actual coding. While there are a few examples of simple all-or-none items, sense organs generally code the logarithm of intensity into frequency. They can do this for any quantity—brightness, loudness, length, and so forth—or for the first derivative, and sometimes for the second derivative. There are also a few that emit impulses at a given frequency when there is no signal and produce a simple increase or decrease of frequency for the strength of signal. For example, the vestibule, the organ of balance and acceleration in the ear, sends in a basic frequency of, say, 39 impulses per second and adds or subtracts from that frequency as the head is accelerated.

Now we return to the problem of the synapse. Some neurons are simple coincidence detectors; some, on receipt of appropriate information, emit a train of impulses; and still others, like those of the balance organs and the big output cells of the cerebellum, deviate in either direction from a fixed resting rate according to the signals reaching them. One of the most instructive groups of cells is the next relay in the vestibular system, for it emits a signal at the frequency which is exactly the difference between the carrier frequency—say 39 per second—which it had been receiving and the frequency that it now receives. So for 36 or 43 it emits 3 per second. In short, for a synapse in general there is no way of knowing what sort of a function the output is of the input. Until you examine it, all you can say is that the synapse is some sort of filter—almost certainly nonlinear—which only says it cannot operate on future information.

The Complexity of Biological Computers*

H. QUASTLER†

***Summary*—Man can handle small amounts of information in conscious purposeful activities. However, the concept of information processing is not limited to these; it applies equally to unconscious nervous control, and even to chemical control and coordination of metabolism. On this level, the amounts of information computed are enormous; man processes about 3×10^{24} bits per day just in the course of producing biochemical tools. The general pattern for these computations is laid down in a blueprint containing 10^6 to 10^9 bits of nonredundant information. Such numbers are characteristic of living things in general. These informational feats are performed with high over-all reliability in spite of frequently low precision of single acts. This is accomplished by prodigious degrees of redundancy.**

MAN'S capabilities of handling information appear sharply limited. In a single act of information processing he can operate the equivalent of about seven independent channels, getting a few bits of information through each. Such acts can occur at rates of up to about 20 seconds. At high speeds, the amount of information per act decreases; a total rate of about 25 bits per second is all that man can sustain, and *that* only by means of highly trained activities and only for limited periods of time.[1] However, these figures do not tell the whole story; they refer only to that fraction of human information processing which can be utilized for transmitting information according to rules imposed from without. They measure man's ability of recognizing particular stimuli in real situations and responding to them by consciously initiating particular actions according to some agreed-upon scheme.

Concurrently with such activities, a vast variety of information processing activities are going on which are not included in the human engineer's measure of performance. For instance, while a man is reading data from some display, he also measures the amount of light which reaches the retina and adjusts the width of his pupils accordingly: this involves processing of information, although he is not aware of it. At the same time, the skin temperature is measured and this information is transduced into orders to blood vessels and sweat glands: that is information processing. The blood is checked for its CO_2 content, and this information serves to regulate respiratory activity: that is information processing. Nor is all interaction with the surroundings restricted to the nervous system. Along the intestinal tract amount and composition of ingested food are recorded and this information is transduced into a spectrum of chemical messages which regulate the secretion of digestive juices: that, too, is information processing. Those juices themselves contain enzymes which sense various chemical configurations and respond to the proper signal with formation of an activated complex and subsequent splitting: even that implies information

* Manuscript received by the PGEC, May 14, 1957. Research carried out at Brookhaven Natl. Lab. under the auspices of the U. S. Atomic Energy Commission.

† Dept. of Biology, Brookhaven Natl. Lab., Upton, N. Y.

[1] G. A. Miller, "The magical number 7 ± 2: some limits on our capacity for processing information," *Psychol. Rev.*, vol. 63, pp. 81–97; 1956.

H. Quastler, "Studies of human channel capacity," in "Information Theory, 3rd London Symposium" (C. Cherry, ed.), Butterworth, Ltd., London, Eng.; 1956.

processing. It is seen that conscious transmitting of information is only a very small part of man's communicating with his surroundings. Furthermore, for any given biological unit such as an individual cell, the vast majority of signals comes not from the outside world but from other cells. Accordingly, interactions with the outside world account for only a very small fraction of the total information traffic and are embedded into a tremendous volume of internal communications.

In considering the complexity of biological computers, it is the enormous amount of total information traffic which is of interest, not the small amount which can be utilized for purposes of conscious information transmission. The use of terms like "information" and "communication" in dealing with involuntary processes is not a matter of loose analogy. There is nothing in the information concepts which restricts their use to voluntary processes. The measure of information is defined by the degree of selection resulting from a decision process and does not depend on why, how, or for what purpose a selection was made.[2] Accordingly, the selection of one out of 20 amino acids by a constellation of chemical forces, based upon the configuration of a reactive surface, is just as good an act of information processing as the selection of one out of 26 letters, initiated by conscious choice, motivated by some idea one man is trying to communicate to another. The amount of information processed depends only on the range of selections available: selecting one out of 20 amino acids is worth up to 4.3 bits, one out of 26 letters up to 4.7 bits, one out of 4 nucleotides up to 2 bits, the upper limit applying when the selection is not prejudiced by unequal *a priori* probabilities.[3]

Most or all biological "decisions" depend ultimately on specificity of chemical reactions. In most or all biochemical processes decisive roles are played by large macromolecules which as a rule have the following characteristics: they have a backbone which is a unidimensional crystal usually of helical shape as linear crystals are bound to be;[4] attached to the backbone at regular intervals are simple organic molecules which form a nonperiodic sequence. These are drawn from pools with a limited inventory. The structural and functional specificity of such macromolecules determines in what acts of information processing they can participate, specifying a macromolecule is equivalent to selecting a particular message out of a set. It appears that this specificity is largely (though not exclusively) determined by the sequence of simple organic molecules lined up along the backbone.[5] The principle is exactly the same as a person's producing a message by means of a typewriter. The movements of the carriage provide the "backbone" of regularly spaced sites; at each site, any one of 83 elements can be placed. The choice of a string of elements specifies the message and the desk, chair, paper, and ribbon are appropriate ingredients of the process. In the same way, the general ambient of the cell interior provides the proper background for coding information in form of molecular sequences. For instance, a specific protein will consist of a few hundred amino acids in a particular sequence. To make this particular protein what it is, amino acids have to be separated from other cell constituents; each amino acid has to be selected out of a set of about 20 and laid down in the proper position. This is done by a chain of processes of selection, each mediated by chemical attraction and repulsion, and somewhere in the cell must exist the information needed to execute the proper selections. This information is coded into a different kind of macromolecule, carrying inscriptions with an "alphabet" consisting of only 4 "letters"; that is, the blueprint of a protein as it exists in a cell looks quite different from the actual protein.[6]

The number of macromolecules synthesized by a man in a single day necessitates the proper arrangement of about one mole (6×10^{23} molecules) of simple organic molecules. Allotting about 5 bits per molecule to make the proper selection (which estimate is probably correct within an order of magnitude), the informational performance is estimated at about 3×10^{24} bits per day, or about 4×10^{19} bits per second. This is a very impressive amount indeed if one considers that a substantial book contains about 10^7 bits of information. Yet the synthesis of specific macromolecules represents only a fraction of the total informational performance of living things and accounts for some of the information needed in constructing metabolic tools, not for the information needed in using them.

Most of the information flow of some 10^{19} bits per second is redundant (as is much of the 10^7 bits in a substantial book). The coordination of redundant information depends on a common pattern. This common pattern is laid down in a network of central stations (cell nuclei, about 10^{14} of them) which were all derived with not very large variations from a single original blueprint. The original blueprint is contained in every fertilized egg, and we are fairly certain that its major portion is coded upon a particular backbone by means of a 4-letter alphabet.[6] There are roughly 10^{10} such letters in a fertilized egg; of these, many appear to be redundant. It has been estimated that the nonredundant information content of the blueprint in the fertilized egg is between 10^5 and 10^9 bits.[3] The information content of the blueprint measures the minimum complexity of the finished structure; the final total information content will be larger because of numerous replications with partial modifications of some replicates based on individual experience.

[2] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication," Univ. of Illinois Press, Urbana, Ill.; 1949.
S. Goldman, "Information Theory," Prentice-Hall, Inc., New York, N. Y.; 1953.

[3] H. Quastler (ed.), "Information Theory in Biology," Univ. of Illinois Press, Urbana, Ill.; 1953.

[4] H. Crane, "Principles and problems of biological growth," *Sci. Monthly*, vol. 70, pp. 376–389; 1950.

[5] L. G. Augenstine, "Protein structure and information content," in "Information Theory" in Biology (H. P. Yockey, ed.), Pergamon Press, London, Eng.; 1958.

[6] G. Gamov, A. Rich, and M. Ycas, "The problem of information transfer from the nucleic acids to proteins," *Adv. Biol. and Med. Phys.*, vol. 4, pp. 23–68; 1956.

On the blueprint level man is only one or two orders of magnitude more complex than a lowly bacterium, but even the blueprint of a bacterium is of very respectable complexity if compared to the information needed to describe efficiently and without repetition a so-called giant electronic brain.

These blueprints of the organism are reproduced by man and bacteria alike with great accuracy and speed. A bacterium will manufacture its own complete blueprint in about 1000 seconds with effective error rates somewhere around 10^{-10} per bit. This is an effective error rate; its low value does not necessarily reflect the precision of the primary acts but could be the result of secondary error checking and correcting. While the blueprint is reproduced, amounts of information several times larger are invested into building and using metabolic tools. In this case there is good evidence indicating that the precision of the individual acts is not high.

It appears to be generally true for most biological systems that the requirements for precision in each single act are low. Yet, although the total amounts of information processed are enormous, the over-all reliability is exceedingly high. Man and animals go through many satisfactorily accurate duty cycles before finally breaking down. This over-all reliability in spite of low elementary reliability must depend on a very high degree of redundancy, although admittedly, we do not know in any detail the codes which are being used for error checking and control.

It is possible that there is a message for engineers in the organization of biological computers. Traditionally, the engineer's ideal is to eliminate redundancy and to perfect elementary reliability, and this is achieved during a more or less extended period of collecting case histories and correcting specific shortcomings. As systems get very complicated, this method becomes less and less feasible. Where there are very many possible shortcomings, it takes too long to find and correct them one by one. Some of the most complicated information handling devices now in use seem to crowd the limits of engineering possibilities. It is possible that such bounds may be broken by using enormous degrees of redundancy to cope with unforeseen as well as with known difficulties; this can be combined with low elementary reliability in order to allow adequate statistical treatment of errors. This can be done if we can learn to rely on design rather than elementary perfection for over-all reliability. These thoughts are heretical but not new. They have been talked and written about[7] but translations into practice have been few and rather half-hearted. The capabilities of radically redundant systems represent a definite challenge to engineering.

[7] *E.g.*, L. N. Ridenour (unpublished).

A Note on the Remarkable Memory of Man*

GEORGE A. MILLER†

THE number of bits per second that a man can take in through his receptors and pass on immediately to another man is quite small; our best estimate at present is about 25 bits per second, and this is obtained only under optimal conditions. When we consider the amount of information a man can hold in mind at one time, as measured by his span of immediate memory, our estimates depend critically upon the way the information is encoded. Most of us can apprehend about five or ten chunks of information at once, where a chunk may contain as much as, say 15 bits—a total of perhaps 100 or 150 bits that we can consider at a time.

In spite of these relatively low limits on the amount of information the average man can actively process, the opinion is often advanced that we have a remarkably large memory, even though it often takes us a long time to retrieve what we know from the store. For example, it has been said that it would require about 1500 bits to store the multiplication table and that the average man must know at least 1000 items as complex as that. As a lower bound, therefore, the human memory must hold at least 1.5 million bits. As an upper bound, we could assume that a person who can process 25 bits per second might remember everything for 16 hours a day for 80 years which gives about 43 billion bits. A middle value of 100 million bits has been called a "reasonable guess."

It is extremely difficult for a psychologist to decide whether such estimates are meaningful. It seems very possible that when we try to recall a fact from memory, what we do is often more analogous to reconstruction than to simple retrieval. For example, even though we are supposed to have memorized the multiplication table, many of us use various devices for inferring, rather than remembering, the larger products (*e.g.*,

* Manuscript received by the PGEC, May 14, 1957.
† Harvard University, Cambridge, Mass.

$7\times 9=7\times[10-1]=70-7=63$). Psychoanalysts have described the reconstruction that goes on in remembering and reporting dreams, and lawyers are quite familiar with the distortions that reconstruction can introduce into testimony. Introspections and anecdotes are not reliable evidence for inferring the capacity of the memory since the things remembered are isolated instances and seldom is their accuracy verified.

An important feature of human cognition is our constant tendency to relieve the strain on our memory by finding rules, or systems of rules, which regenerate the facts of our experience but are shorter and easier to store than the experiences which they represent. For example, on the basis of a few, perhaps a hundred, grammatical rules that we acquire after a finite exposure to our native tongue, we are prepared to generate or to understand any one of an infinite set of grammatical sentences. Or again, on the basis of a simple formula and the rules of mathematics, we are prepared to summarize the results of thousands of experimental observations on falling bodies, or the volume of a gas, or the positions of the stars, etc. Or, on the basis of simple classifications like "Jew," "Negro," or "Nazi," we are prepared to behave in a uniform way toward thousands of our fellow men. We usually reorganize and simplify our experience before we try to store it away in memory.

We are not always consciously aware of applying these rules. Thus, it is extremely difficult to tell from introspection whether an item was recalled directly from memory or was calculated by rules that are stored in memory. If, as seems likely in most cases, the answer is calculated rather than stored, we cannot estimate the amount of information a man has stored until we know the rules of calculation that he used.

And, the rules cannot be observed directly in most common instances.

At the present time, therefore, it is more important to determine the kinds of rules we use than to put numerical estimates on the available size of the human memory. It may yet be possible to account for the remarkable memory of man with fewer bits of storage than we currently estimate.

The Human Computer in Flight Control*

LAWRENCE J. FOGEL†

Summary—**It is the unique decision-making ability which makes the human operator particularly suited to accomplish flight control computation. His decisions range from the simplest repetitive determination of control actions, each intended to minimize the error, to complex problems of the sequential reprogramming of all the lower levels of decision which, in totality, comprise his information transduction process. The human sensory-input and motor-output channels must be properly used in order to insure maximum efficiency of the over-all human computer. The perceived noise can only be separated from the signal after the decision-making operation to be performed has been clearly identified. Human noise filter and information handling rate characteristics may be enhanced through the utilization of inherent qualities of the various input and output channels.**

A brief summary is made of the considerable effort which has been devoted toward the development of mathematical models for human tracking. It is emphasized that the human decision-making ability must be incorporated into the model before a realistic representation can be attained. A stochastic model is considered which is constructed in a heuristic manner, incorporating the decision-making ability. Specific direction is indicated which may yield the most fruitful approach to the mathematical analysis of human flight control computation. This future model would be of great value as a link between the related work in the fields of neurophysiology, biophysics, and psychology. Most important, it should provide a cockpit synthesis technique which will aid the design of future manned aircraft.

* Manuscript received by the PGEC, May 14, 1957.
† Convair, Div. of Gen. Dynamics Corp., San Diego, Calif.

Introduction

THE unique ability which makes the human operator particularly suited to the performance of flight control is that he can program and reprogram his computation while the flight is in progress. The word "computation" is intended in its broadest sense: the transduction of the received data into corrective control actions to accomplish the desired mission.

This reprogramming may be looked upon as a decision which occurs at a particular level within a hierarchy of decisions relevant to the mission. This hierarchy may be grossly characterized as follows:

1) Decision that determines the general mission program to accomplish the flight objective,
2) Decision that determines the implicit flight path to accomplish this desired program,
3) Decision that determines the computation program which will be used to process the data so as to accomplish this flight path in an optimal manner,
4) Decision that determines the state of the aircraft from the displayed information,
5) Decision that determines the manner of control operation which will minimize some function of the anticipated error.

These classifications are not mutually exclusive, but it is easy to see that they are distinctly ordered. They must be performed in an ordered sequence to lead to effective flight control.

It is obvious that the highest level decision must be accomplished in accordance with the requirements of air traffic control or the assigned strategy from military headquarters. The pilot must initiate his activities to execute the results of such decisions. He must be aware of their outcome but he does not take part in the decision-making at this level. The remaining levels of decision are normally left to the pilot who pre-plans the gross flight path (and files a flight plan) and chooses the detailed airpath while the flight is in progress. His flight plan is a result of the application of a set of procedures and constraints to the presented problem of mission, traffic, weather, etc. His airpath selection is accomplished by reference to his experience with weather, fuel consumption, flight time, and other characteristics of the aircraft. For example, an interceptor pilot may preplan a minimum time climb to the target altitude; however, as the situation develops he may modify this chosen airpath so as to approximate minimum fuel consumption requirements, or be of some intermediate nature between these two extremes. The pilot must rely upon his memory, the summation of his experience to accomplish this latter decision.

The remaining levels of decision are the subject of the following discussion. The pilot's decision at the third level is concerned with the program for his information transduction processing. The fourth level relates to the decisions which establish the nature of the data received for processing, the separation of signals from noise. The fifth level decisions pertain to the required modification of the output data to allow efficient coupling to the aircraft system.

The human computer offers remarkable flexibility but also certain specific limitations. The fullest use can only be made of the man if his characteristics are well known to the cockpit designer. It is the purpose of this paper to examine some of the most important flight control qualities of human information transduction.

Discussion

Prior to considering the information transduction process, it appears desirable to examine some aspects of the human computer's input and output channels. The battery of senses with which man is equipped provides a mass of data about his environment. Unfortunately, much of this information is noise and very little is signal; noise being defined as any information not of direct relevance to the decision which he is about to make. It, therefore, follows that the nature of the decision must be determined before any judgment may be reached regarding what is signal and what is noise. To illustrate, consider the pilot's reaction to an extra-tight shoulder harness. If his immediate concern is fuel management, these data received through his tactual sense are a bothersome disturbance, a noise which can only degrade his decision-making ability. On the other hand, if his concern happens to be the adjustment of the shoulder harness, the degree of this tightness is the information he desires to examine.

Noise may be either spatial or temporal, or both, dependent upon the sensory channel which is used to receive the information and the nature of the sequential decisions. Such disturbance can only be minimized by providing the pilot with a tight coupling communication to and from the aircraft system. Such a coupling is characterized by the term "compatible." Before considering some details of the actual information transduction, it seems worthwhile to note that the importance of compatibility is as great in using the human in flight control as is making the proper input and output connections on any laboratory computing device. In the latter case, an error is quickly recognized and corrected; however, coupling errors between the man and machine are often difficult to detect and, once detected, often are almost impossible to completely eliminate.

Appendix I indicates some of the man vs machine characteristics which should form a guide to the choice of the type of function to be left to the pilot. Assuming that this primary operational design has been determined, compatibility may be increased by noting that each human sensory channel has individual time adaptation and importance characteristics, as well as a maximum number of degrees of freedom. For example, the kinesthetic channel has adaptation which is dependent upon the imposed "g" force vector of a maneuver. Human sensing of the vertical to the earth should note this time relaxation characteristic of the kinesthetic display which *always* feeds information to the pilot. To ask the pilot to operate from one visual "vertical" while his body displays another "vertical" is imposing a condition which might easily induce vertigo.

Although the tactual and olfactory channels are severely limited as far as precision and channel capacity are concerned, they are the most natural channels for the display of warning or emergency status. The most natural meaning associated with a sudden tactual stimulation is "danger." Such a perceived stimulus cannot be ignored by the pilot, as he might easily miss a warning light or even fail to interpret properly a klaxon horn which tries to inform him of an impending "wheels up" landing. It is through these channels that low probability, high surprise value information events are the most efficient information to display.

Each mode of display has a "dimensionality"; the number of independent qualities which can be used to portray additional information. For example, in a two-dimensional visual display a point position corresponds to two coordinate data. In addition, the position may be coded with a shaped block. Shape then becomes another dimension of the display. Color is another possibility, etc. These dimensions are not all usable with the same precision or accuracy but each can contribute some additional information, provided the human channel capacity is not exceeded. Probably the simplest and most efficient visual display dimension is the position of a point along a coordinate axis. Measurements have

shown the amount of average information which can thus be identified to be in the order of 3 bits per dimension.[1,2] The use of two points on two scales almost doubles the visual display information capability. Three scales offer some added information but the sum of the three information rates is certainly not the same as three times each scale presentation. Further increase of dimensionality soon reaches a crossover point at about eight or nine coordinates. Any further increase degrades the visual display.[3]

Certainly, it is valuable to supply all the required information to the pilot but such display takes its toll in terms of fatigue,[4] inaccuracy in his information processing, and the imposition of a stressful condition which makes the operator introduce subjective noise and clutter into the displayed data between reception and perception.[5] When the human operator functions near channel capacity,[6–8] it is normal to find that each error sets off a train of succeeding errors due to the additional data provided by his recognition of the first error. The channel capacity has been exceeded and the excess information offers confusion which in turn decreases the remaining channel capacity.[9,10] The human operator has wonderful ability to extrapolate data[11,12] and make *a priori* decisions. This characteristic will be analyzed in detail below, but it is only relevant at this point to recognize that this device provides an information "smoothing" which can allow a little forethought to keep the pilot from being pushed beyond his limiting channel capacity.

In dealing with the human operator in a multivariate situation, it serves to increase efficiency as well as flight safety to allow each variable to take on a specific designatory meaning. It yet remains to determine an analytic technique for measuring the amount of this meaning,[13] however, meaning certainly aids in the separation of parameters, elimination of possible ambiguity errors, and memory access. To maximize the benefit from this human ability, it is well to offer an over-all cockpit structure for both displays and controls which provides a contextual background for the acceptance of data from each flight parameter. To illustrate, consider the logic of a proposed cockpit which provides pictorial information about attitude and map position in the center, then proceed to the right to provide the result of angle of attack and pitch—rate of climb. It is followed, reading further to the right, by the resulting altitude which may be read in numeric value from left to right with succeeding levels of increased precision. The right hand has access to the pitch, angle of attack, and roll attitude control.

Proceeding to the left from the central attitude and map display first provides forward motion data, angle of attack, indicated air speed and Mach number. Velocity information leads to concern for the passage of time—a numerical clock is placed over the engine instruments which supply the data concerning the causal thrust. These engine instruments, themselves, should be arranged to agree with the causal relationship so that fuel gauges would reside at the extreme left of the instrument panel being directly adjacent to the throttle control which is operated by the pilot's left hand. Such an over-all logic seems to offer a rationale for aircraft panel standardization which provides a positional context to assist the identification and perception of the many variables.

The efficiency of the output coupling of man-to-machine is a function of the control stick sensitivity and displacement range. Standard human engineering data have appeared which describe efficient designs in this regard.[14] Obviously, the coupling of binary decisions is better suited to switch or push-button devices, while continuous variables are best communicated to the machine by position or force sensing control sticks. The configuration of these controls should directly relate to the display configuration so as to minimize any coordinate transformation which would be required on the data resulting from human transduction.[15,16]

The often suggested "ultimate" in control design should be re-evaluated. It consisted of manual controls which agreed with the earth-reference coordinate system. The pilot's over-all purpose would be simply executed by a vertical movement of a control member to change the aircraft altitude and a horizontal movement

[1] I. Pollack and E. T. Klemmer, "The Assimilation of Visual Information from Linear Dot Patterns," AFCRC Tech. Rep. 54-16, Air Force Cambridge Res. Center; July, 1954.

[2] E. T. Klemmer and F. C. Frick, "Assimilation of information from dot and matrix patterns," *J. Exper. Psychol.*, vol. 45; November, 1953.

[3] J. C. R. Licklider (ed.), "Problems in Human Communication and Control," paraphrased transcription of a conference sponsored by the Natl. Sci. Found., Mass. Inst. Tech., Cambridge, Mass., pp. 15–17; 26; June, 1954.

[4] J. P. Egan and E. J. Thwing, "Further studies in prestimulatory fatigue," reprint from *J. Acous. Soc. Amer.*, vol. 27, pp. 1225–1226; November, 1955.

[5] J. Deese and R. Lazarus, "The Effects of Psychological Stress Upon Perceptual-Motor Performance," Res. Bull. 52-19, Air Force Human Resources Res. Center, Lackland AFB, San Antonio, Texas.

[6] E. R. F. W. Crossman, "The information-capacity of the human operator in symbolic and non-symbolic control processes," in "The Application of Information Theory to Human Operator Problems," W. R. D. Rep. No. 2/56 Directorate of Weapons Res., Proc. Special Tech. Meeting at Royal Empire Soc.; September 19, 1955.

[7] "Human Performance in Information Transmission," Rep. R-62, Control Sys. Lab., Univ. of Ill., Urbana, Ill.; March, 1955.

[8] "Human Performance in Information Transmission. Part III—Flash Recognition-Scale Reading," Rep. R-68, Control Sys. Lab., Univ. of Ill., Urbana, Ill.; October, 1955.

[9] L. S. Christie and R. D. Luce, "Suggestions for the Analysis of Reaction Times and Simple Choice Behavior," Rep. R-53, Control Sys. Lab., Univ. of Ill., Urbana, Ill.; April, 1954.

[10] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *J. Exper. Psychol.*, vol. 47; June, 1954.

[11] R. M. Gottsdanker, "The continuation of tapping sequences," *J. Psychol.*, vol. 37, pp. 123–132; 1954.

[12] E. C. Poulton, "Speed Anticipation and Course Anticipation in Tracking," APU 123/50, Medical Res. Council, Cambridge, Eng.; 1950.

[13] L. J. Fogel, "Toward a measure for meaning," Proc. IRE, vol. 43, p. 1018; August, 1955.

[14] "Handbook of Human Engineering Data," Human Eng. Rep. SDC 119-1 (Nav. Exos P-643), 2nd ed., Tufts College Inst. for Applied Experimental Psychology, Medford, Mass.; December, 1951.

[15] R. L. Deininger and P. M. Fitts, "Stimulus-response compatibility, information theory, and perceptual-motor performance," in "Information Theory in Psychology," by H. Quastler, The Free Press, Glencoe, Ill.; 1955.

[16] P. M. Fitts and C. M. Seeger, "S-R compatibility: spatial characteristics of stimulus and response codes," *J. Exper. Psychol.*, vol. 46; September, 1953.

to alter its heading; these actions being transformed into attitude directives by a computer. No consideration has been given to the compatibility of such controls with the kinesthetic display sensed by the pilot. So long as manned aircraft have wings, it appears of greatest value to allow the pilot to directly manipulate the attitude of the aircraft, and in this manner achieve the highest degree of compatibility between the control and the display.

The human information transduction itself has been the subject of a large number of investigations with the consistent effort to express the transfer characteristics in some mathematical form.[17–21] Much has been accomplished, yet the universally recognized nonlinearity and time varying characteristics make the task extremely difficult. The initial efforts looked toward approximating a linear invariant transfer function for various types of stimuli including sudden step functions of different amplitude, random continuous motion of known bandwidth, and periodic waveforms. The frequency response of the tracking was plotted on a Bode diagram and fitted in standard servo analysis manner (using straight line db vs frequency fall-off characteristics from the identified break points) to provide the form of the transfer function. McRuer has done an excellent survey of this work[22] and has summarized the findings in the form of a control to visual displacement ratio,

$$Y_p = \frac{K_p e^{-\tau s}(T_L s + 1)}{(T_I s + 1)(T_N s + 1)}$$

where:

1) The reaction time delay is represented by $e^{-\tau s}$ with τ having values ranging from 0.2 to 0.5 second for average random stimuli. If the stimulus becomes predictable to the pilot, he may begin to generate an output which replicates the input and is synchronized with it. When such is the case, τ becomes negligible. Any phase discrepancy is not due to the reaction time delay. On the other hand, τ may be greater than 0.5 second dependent upon the interpretability of the perceived data.
2) The neuromuscular lag, T_N, is normally within the range between 0.1 and 0.16 second for the arm.
3) The lead time constant, T_L, has been observed to have values between 0.25 and 2.5 seconds; however, these values are certainly not the limits to its range. This constant is a function of both the dynamic response of the aircraft and the bandwidth of the visual stimulus. This linear factor in the numerator provides a 6 db/octave rise in the gain characteristic from the break point identified by $\omega = 1/T_L$ which may be looked upon as the added importance the higher frequency harmonics receive as they infer anticipatory information.
4) The lag time constant, T_I, with observed values between 5 and 20 seconds can have any value, dependent upon the aircraft dynamics and the stimulus bandwidth. This "integrating" linear factor provides a smoothing of the input data so as to allow the output spectrum generated to approximate better the response spectrum of the aircraft. The closer this term approximates pure integration, the more relative importance the pilot has attributed to "drift components" of the stimulus.
5) The gain, K_P, is adjusted by the pilot to allow proximity to the point of marginal stability. For tasks requiring greater sensitivity and accuracy he would raise the gain. For more relaxed flying "on a Sunday afternoon," he would decrease the gain and allow the aircraft to "fly itself" more. The accuracy criterion which appears most likely[23] is the minimization of the rms error. This criterion is also dependent upon the presumption of time invariant transfer linearity.

A number of nonlinear models has been considered. One report[24] indicates the additional assumptions of rate judgment (by adding a weighted derivative of the stimulus), rate threshold (which notes that certain observed errors are normally disregarded as being of negligible importance since their average value of zero will be approached in a relatively short time), and clamping (a first order hold which the pilot normally applies to control between the correction of new errors). This model was a sufficiently close approximation to allow its substitution for the pilot in a cruise simulation for periods up to thirty seconds without the pilot's detection of the substitution. The particular values chosen for this "pilot analog model" were taken from the residue error from previous analyses of tracking control using the aforementioned linear invariant model.

An alternate nonlinear approach was suggested by J. Elkind. It rested upon choosing a family of linear transfer functions which were interchanged depending upon the amplitude of probabilist aspects of the stimulus. Another presumed to find a piecewise linear fit best as a function of amplitude based on empirical data plotted on the phase plane.[25] Only cursory experiments

[17] H. P. Birmingham and F. V. Taylor, "A Human Engineering Approach to the Design of Man-Operated Continuous Control Systems," NRL Rep. 4333, Naval Res. Lab.; April 7, 1954.

[18] C. E. Walston and C. E. Warren, "A Mathematical Analysis of the Human Operator in a Closed-Loop Control System," Res. Bull. AFPTRC-Tr-54-96 Air Force Personnel and Training Res. Center, Lackland AFB, San Antonio, Texas; 1954.

[19] C. E. Walston and C. E. Warren, "Analysis of the Human Operator in a Closed-Loop System," Res. Bull. 53-32, Air Force Human Resources Res. Center, Lackland AFB, San Antonio, Texas; August, 1953.

[20] "The Human Pilot," BuAer Rep. AE-61-4 III, Bureau of Aeronautics, Navy Dept., Washington, D. C.; August, 1954.

[21] J. R. Ragazzini, "Engineering Aspects of the Human Being as a Servomechanism," unpublished paper presented at the 1948 meeting of the Amer. Psychol. Assoc.

[22] D. McRuer, "Dynamic Response of Human Operators," Memo. Rep. No. 53, Control Specialists, Inc., Los Angeles, Calif.; November 10, 1955.

[23] J. I. Elkind, "Characteristics of Simple Manual Control Systems," Tech. Rep. No. 111, Lincoln Lab. M.I.T., Cambridge, Mass.; April 6, 1956.

[24] "Final Report—Investigation of Vestibular and Body Reactions to the Dynamic Response of a Human Operator," Goodyear Aircraft Rep. GER-5452; November, 1953.

[25] L. J. Fogel and J. Senders, "The Human Operator as a Multi-Mode Servomechanism," presented at the Conference on Man-Machine Relations, The Franklin Inst., Philadelphia, Pa.; February, 1955.

were performed but the separation of domains and switching lines appeared possible. Other investigators were concerned with the hypothesis of a sampled data servosystem model;[26,27] however, as yet no sufficiently complete study has been performed which would prove the hypothesis that the pilot actually operates only at discrete times. Of course, application of the generalized sampling theorem[28] makes it possible to show that certain empirical data are equally valid for a continuous or discrete time model.

The essential stochastic nature of the human operator was recognized by J. North who reported a statistically variant linear model to account for the variability and pilot error in tracking. In references[29–31] he states, "Some psychologists have considered that the stochastic oscillations were evidence of a discontinuous process; this is seen not to be the case." His point is well taken; however, no conclusive proof has yet been presented which will confirm or deny this long standing hypothesis.

The purpose of any mathematical model is to allow the prediction of empirical data; however, this is not quite enough. The model should also be based on some heuristic foundation, so that an hypothesis concerning the subject of investigation can be turned into a principle or disregarded in favor of some new hypothesis. There is required some practical compromise in the formation of such a testable hypothesis. First, it must be a reasonable guess concerning the nature of the human operation. This could easily lead to forbidding complexity. On the other hand, the conceptual model should be simple enough to allow for mathematical computation. The greatest advance is made through the use of the most specific conceptual model which is mathematically expressible in complete form and can be validated by empirical data.

A recent paper[32] proposes a purely stochastic model for the tracking behavior of a pilot. This model is different in that it speculates the single-valued operation of the human so that the pilot forms a sampled data system for each of the many parameters he considers. First order hold smoothing is utilized upon each discrete error correction control to synthesize the actual continuous multiparameter operation of the pilot. The model is expressed as a cascade of operations, each defined by the transformation of the probability density function associated with the received signal element. Such a model allows human anticipation to be expressed in the reasonable form of a diffusing probability distribution which agrees with the known informational aspects of human prediction. The sampled data operation requires the use of a memory of at least the last data for each parameter. This memory operation satisfies the same specific hypotheses as the anticipator element once the time origin has been shifted from the present to that time in the past when the memory received the information of immediate concern. Fig. 1 illustrates this model and Appendix II indicates that portion of the mathematics which describes the anticipator-memory operation.

Currently an experimental program is in progress at Convair for the purpose of obtaining design data for optional anticipatory displays. A simulated aircraft is controlled by naive subjects and jet pilots with the automatic recording of the error criteria as well as the Galvance skin response of the subject. Cursory trials were based on previous work[33–35] and seem to indicate that this latter measure identifies the amount of anticipation when a single discrete time event is about to occur. These same data can also be interpreted for other information about the subject, such as stress level, level of alertness, fatigue, etc.

A "rules of choice" element is included which presumes a time-varying transfer characteristic which expresses the weighting of importance the operator imposes on the perceived magnitude of the signal element. Here is included both the operational threshold which is chosen in accord with the required precision of flight and the saturation quality the importance weighting must have as it reaches a maximum magnitude. Modification of this transfer characteristic is what has been previously referred to as re-programming, as one corresponding to a variational calculus problem.

Tracking may be looked upon as a decision process where the same decision to minimize some function of the error is sequentially repeated. The error criteria may be taken as invariant provided the higher level reprogramming decision is also allowed. In fact, the same functional criteria could be the basis for the choice of the importance weighting program itself. Such speculation emphasizes that the one feature which has been omitted from previous models is probably the *most* essential element; that is, the decision-making ability.

Decision theory has only recently received intensive investigation;[36,37] however, most attention has been

[26] W. E. Hick, "The discontinuous functioning of the human operator in pursuit tasks," *Quart. J. Exper. Psychol.*, vol. 1, pp. 36–51; 1948.

[27] J. M. Stroud, "The fine structure of psychological time," in "Information Theory in Psychology," (H. Quastler, ed.), The Free Press, Glencoe, Ill.; 1955.

[28] D. Jagerman and L. J. Fogel, "On some general aspects of the sampling theorem," IRE Trans., vol. IT-2, pp. 139–145; December, 1956.

[29] F. C. Bartlett, "The Measurement of Human Skill," Parts I, II, *Brit. Medical J.*, pp. 835–838; June, 1947.

[30] J. W. Craik, "Theory of the Human Operator in Control Systems," Parts I, II, *Brit. J. Psychol.*, vol. 38, pp. 56–61; 1948, and vol. 38, pp. 142–148; 1948.

[31] M. A. Vince, "The intermittency of control movements and the psychological refractory period," *Brit. J. Psychol.*, vol. 38, pp. 149–157; March, 1948.

[32] L. J. Fogel, "An analysis for human flight control," 1956 IRE Convention Record, Part 8, pp. 69–88.

[33] E. A. Haggard, "Experimental studies in affective processes: I. Some effects of cognitive structure and active participation on certain autonomic reactions during and following experimentally induced stress," *J. Exper. Psychol.*, vol. 33, pp. 257–284; 1943.

[34] E. A. Haggard and R. Gerbrands, "An apparatus for the measurement of continuous changes in palmar skin resistance," *J. Exper. Psychol.*, vol. 32, pp. 92–98; 1947.

[35] E. W. Geldreich, "The use of a calibrated potentiometer in the measurement of the galvanic skin response," *Amer. J. Psychol.*, vol. 47, pp. 491–493; 1935.

[36] A. Wald, "Statistical Decision Functions," John Wiley & Sons, Inc., New York, N. Y.; 1950.

[37] D. Blackwell and M. A. Girshick, "Theory of Games and Statistical Decisions," John Wiley & Sons, Inc., New York, N. Y.; 1954.

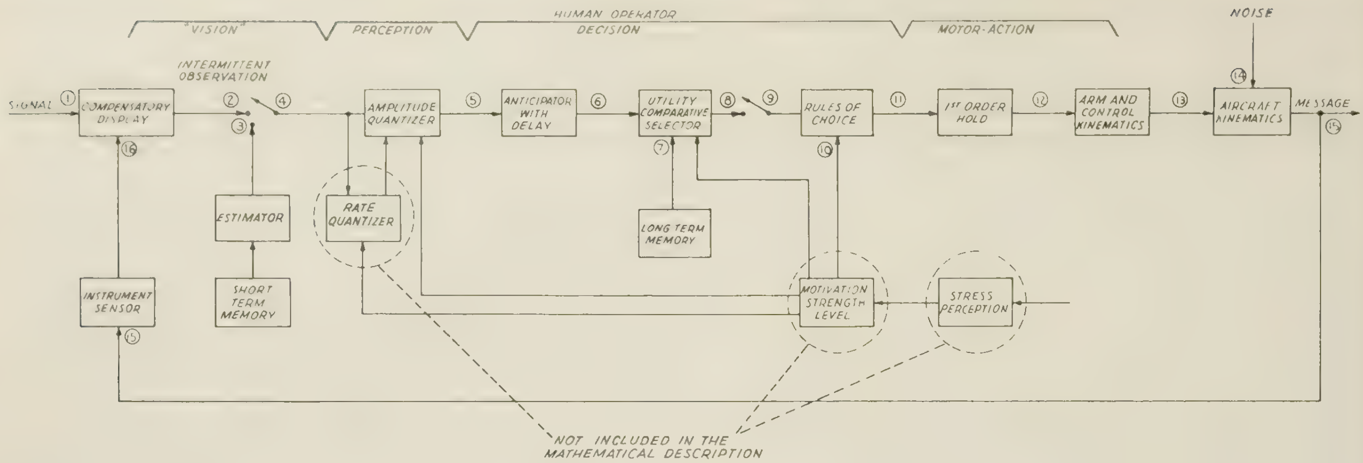


Fig. 1—Single parameter mathematical model of information transfer system.

focused on the mathematics of optimal decision making (the theory of games). The concern here, however, is how the human pilot makes decisions and this can be a very different matter under some circumstances. Several papers[38–40] have noted this disparity and some propose alternative calculi for the model of human decision making.[41,42] This field is young and awaits exhaustive investigation of the individual models. Much effort has been devoted to the detection decision.[43] "Is there a signal embedded in the perceived noise?" This latter investigation emphasizes the subtle nature of human decision making by pointing out that information which is displayed below the level of conscious awareness is still received by the human operator and may be extracted from him by a forced choice situation. This finding removes the traditional belief in a fixed lower threshold to perception and replaces it by a probability of reception which asymptotically approaches zero for decreasing size stimuli.

Conclusion

Man's greatest value as a pilot can only be achieved when his decision-making capacity is utilized to the fullest extent. These decisions range from the simplest choice of error correcting actions to the selection of a transduction computation program. Further studies directed toward structuring a mathematical representation for the human operator should find it possible to form a conceptual model which includes the fundamental decision-making operation. Such a model would serve a dual purpose. It would provide a means for the prediction of human behavior in proposed environments and it will, in itself, become the fusing link to bind together the work which has simultaneously been carried on in neurophysiology[44,45] biophysics,[46–48] and psychology.[49,50]

Methods of analysis can be most fruitful if they attempt to unite the thinking and previous findings in related fields of interest. The present state of the art provides new tools for the fabrication of such a model. If this step can be accomplished, it will offer a distinct advantage to all the contributing fields. Engineering will emerge with a new synthesis technique for cockpit design. One which will optimize the environment for the human decision maker.

Appendix I

Some System Function Attributes of Man as Compared to Machines

The following tabulation will be categorized by system function properties, listing those in which the machine is usually superior first, then considering those in which the human excels. Since these are system functions, the human properties will be predominantly psychological, describing aspects of the human transduction of information. This is not an attempt to be com-

[38] W. Edwards, "The Theory of Decision Making," *Psychol. Bull.*, vol. 51; July, 1954.

[39] W. Edwards, "An Annotated Bibliography of Very Recent Work on Decision Making," unpublished draft, to be submitted to *Econometrica*.

[40] S. Vail, "Alternative calculi of subjective probabilities," in "Decision Processes" (R. M. Thrall, G. H. Combs, and R. L. Davis, eds.); John Wiley & Sons, Inc., New York, N. Y.; 1954.

[41] L. J. Fogel, "On the Design for an Optimal Training Strategy," Convair Rep. ZG-8-024; October 9, 1956.

[42] W. Edwards, "An Empirically Oriented Model for Decision Making," Lab. Note AEPRL-LN-55-10, Air Force Personnel and Training Res. Center, Lowry AFB, Denver, Colo.; October 7, 1955.

[43] W. P. Tanner, Jr., "Psychophysical Application of the Theory of Signal Detectability," Eng. Res. Inst., Univ. Mich., Ann Arbor, Mich.; November 13, 1953.

[44] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5; December, 1943.

[45] L. A. Jeffress (ed.), "Cerebral Mechanisms in Behavior," John Wiley & Sons, Inc., New York, N. Y.; 1951.

[46] N. Rashevsky, "Mathematical Biophysics," The University of Chicago Press, Chicago, Ill.; 1948.

[47] W. R. Ashby, "Design for a Brain," John Wiley & Sons, Inc., New York, N. Y.; 1954.

[48] H. Quastler, "Information Theory in Biology," University of Illinois Press, Urbana, Ill.; 1953.

[49] H. Quastler, "Information Theory in Psychology," The Free Press, Glencoe, Ill.; 1955.

[50] R. R. Bush and F. Mosteller, "Stochastic Models for Learning," John Wiley & Sons, Inc., New York, N. Y.; 1955.

plete or to paint any black and white picture in any dimension; rather, it is the purpose of this Appendix to indicate the type of thinking indigenous to human engineering which maintains a proper perspective in the application of biotechnology to actual hardware systems design.

Machines Excel in the Following Functions

In handling of high inertia components or smaller masses over longer periods of time a machine can be designed to tolerate higher imposed forces than the human. So it is that missiles may be guided through more aerodynamic rigors than could be accomplished with a manned aircraft. Aside from the sheer physical abilities, the machine is superior in the reliability of repetitive functioning. In such operations fatigue is usually only a secondary consideration. The machine does not suffer from "boredom" or other subjective qualities. For example, it remains "unconscious" of its environment and so performs consistently with the programming regardless of imposed "distractions" or dangers to survival and, in general, this is as it should be since the machine is more expendable than the man. By proper design, a machine can be made to have almost immediate response. This could be a distinct advantage over the unremovable human reaction time latency in particular applications. It can transduce information with a channel capacity which can be made arbitrarily large and such large inputs can be effectively stored in a precise but short-term memory which is superior to the human ability. It can utilize these data in the solution of extremely complex problems at rates far in excess of anything a man could accomplish.

Man Excels in the Following Functions

In pattern recognition man excels in the ability to filter "noise" from randomly presented data and identify a signal structure in the given data. In fact, such structure is postulated even when the subject is given truly random pure noise.[18] This ability is common to both the spatial and temporal domains and is dependent upon the remarkable human correlative memory and axes transformation capability. This latter characteristic may be illustrated by human observation of a two-dimensional pattern as representation for a three-dimensional object which can then be mentally "viewed" from any desired direction. The above-referenced time domain filtering also allows for the establishment of a knowledge of the sequential dependencies. This, in itself, is sufficient information to allow of self-reprogramming to account for transient and nonstationary characteristics in the perceived data. In fact, it would prove extremely costly to build a device which is so well suited to the monitoring of low probability events. This is true in any single dimension and becomes even more impressive when the human ability to receive information by any of a diverse set of multimode sensors. These can operate at very great sensitivity. Ordinary equipment does not compare in either the number of available sensory modes or the visual sensitivity which is available; however, it is still possible to build special purpose devices which will surpass the human in these respects provided the cost and time required are available. Some of the channels are particularly suited to the sensing of information which warns of danger and these "handy" channels, although not in direct use, couple to the reprogramming ability to offer great flexibility and resultant reliability in dealing with unexpected circumstance. He can accomplish the same purpose with a different approach. Complex equipments utilize redundant design which increases reliability, but at an exponential cost in weight and number of components. It is also apparent that the human operator is not subject to jamming by any of the ordinary ECM techniques, and in fact his spectral filter characteristics allow his detection of masked signals taken from radar or sonar receivers. Usually the man is required to take some continuous corrective action based upon his acquisition of a target. Here the machine can prove superior in tracking except for the variable weighting which the human operator can apply to displacement error, rate of error, acceleration of error, and other nonlinear properties should they become desirable to stabilize the closed loop. Such nonlinearities prevent system overload and, when it is properly motivated, the human can exhibit extraordinary properties which can serve to compensate for inadequacies of the controlled system. Certain absolute limits cannot, however, be exceeded. For instance, the channel capacity for the transduction of information sets a breakdown point beyond which the human can be pressed. Even within these limits his mental computation is weak and relatively inaccurate. He cannot be expected to follow an optimal theory of games solution, although in any real world complex situation his solution will probably prove closest to that desired optimal since a computed solution cannot always be obtained. To conclude, it may be stated that the human operator is relatively light in weight, and relatively easy to maintain. When given a well-designed cockpit he is in good supply and relatively inexpensive to produce.

Appendix II

The Anticipator-Memory Model

A simple model for the estimator or anticipator can be determined from certain logical hypotheses concerning the expected transformations of the moments. The uncertainty of the output must be some direct function of both the time interval of projection, as well as the uncertainty of the input. Let x, y, and N represent the input, output, and noise variable of the estimator.

Hypotheses:

1) That the mean value of the output random variable, m_y is equal to the mean value of the input variable, m_x.

2) That the variance of the output random variable, σ_y^2, is equal to the variance of the input random variable, σ_x^2, increased by a factor proportional in some manner to the estimation or anticipation time interval, α.
3) That the skewness of the output random variable, η_y, will usually be some decreasing function of the estimation or anticipation time interval, α.
4) That the added noise of estimation or anticipation, N, is independent of the input variable, x, but has the same moments of the input variable with the exception of the mean value which is zero. In this manner, the noise which is added is in itself directly dependent upon the indefiniteness of the input random variable.

These required hypotheses may be fulfilled by a simple time domain model

$$y = x - k\alpha N, \qquad k > 0. \tag{1}$$

In graphic form, this equation defines a circuit as shown in Fig. 2. From this model it may be seen that

$$m_y = m_x - m_{N'} \tag{2}$$

where

$$m_{N'} = k\alpha m_N \tag{3}$$

but, by hypothesis

$$m_N = 0 \tag{4}$$

so that

$$m_y = m_x \tag{5}$$

$$\sigma_y^2 = \sigma_x^2 - 2\alpha_{xN'} + \sigma_{N'}^2. \tag{6}$$

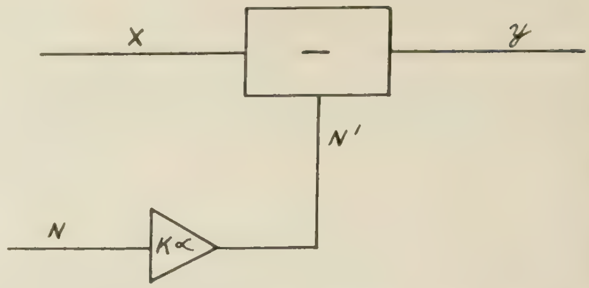


Fig. 2—Time domain model for the estimator of anticipator element.

Due to the hypothesized independence of N and x,

$$\alpha_{xN'} = 0. \tag{7}$$

Further,

$$\sigma_{N'}^2 = k^2\alpha^2\sigma_N^2 = k^2\alpha^2\sigma_x^2, \tag{8}$$

so that

$$\sigma_y^2 = (1 + k^2\alpha^2)\sigma_x^2 \tag{9}$$

$$\eta_y = E[x^3 - 3k\alpha N x^2 + 3k^2\alpha^2 N^2 x - k^3\alpha^3 N^3], \tag{10}$$

but $m_N = 0$ so that

$$\eta_y = \eta_x + 3k^2\alpha^2\sigma_N^2 m_x - k^3\alpha^3\eta_N. \tag{11}$$

So long as

$$\frac{\eta_x}{\sigma_x^2} = \frac{\eta_N}{\sigma_N^2} > 3k^{2/3}\alpha^{2/3}m_x \tag{12}$$

is anticipated, a monotonically decreasing function will result.

The value of this model appears to be its simplicity and its intuitive agreement with some psychophysical evidence which concurs with most of the usually identified aspects of the process of estimation or anticipation

Correction

Velio A. Marsocci, author of "An Error Analysis of Electronic Analog Computers," which appeared on pages 207–212 of the December, 1956 issue of these Transactions has requested the editors to make the following corrections to his paper:

Eq. (42) should read

$$\Delta x_n = -\frac{S_n T_2[S_n^m + \sum B_i S_n^i]}{T_2[mS_n^m + \sum B_i(iS_n)^i] + \left[\dfrac{T_2(S_n^m + \sum B_i S_n^i) + C^1(S_n)}{2T_1 S_n + 1 + \dfrac{T_1}{T_0}}\right]}$$

and the statement following (30) should read,

"... equivalent to saying the $X_{n0} = S_n$."

Correspondence

A Method for Evaluating Amplifier Phase Shift at Low Frequencies*

Consider the circuit configuration using ideal components shown in Fig. 1. The open loop transfer function for this circuit is:

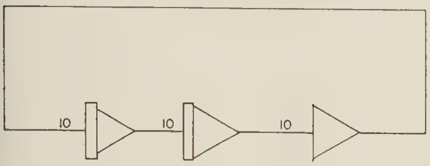


Fig. 1

$$W(S) = -1000/S^2.$$

The characteristic equation for the closed loop system then is:

$$S^2 + 1000 = 0.$$

Examination of the roots of the above function indicates that at $S = j10\sqrt{10}$, the system has an undamped resonant frequency at $\omega = 31.6$ radians.

A practical operational amplifier will have some small but finite phase shift and attenuation at this frequency. Attenuation will result in a change in the ω_r while a small phase shift will primarily introduce an exponential decay or increase in the amplitude of the sinusoid.

If the expression for the open loop transfer function is developed for a practical system of the type above, where integration approaches the ideal (see Appendix), and the summing amplifier has a small lagging phase shift, then

$$W(S) = \frac{-1000}{S^2(1 + \tau S)}.$$

The characteristic of the closed loop system is:

$$S^2(1 + \tau S) + 1000 = 0$$

$$\tau S^3 + S^2 + 1000 = 0$$

$$S^3 + \frac{1}{\tau} S^2 + \frac{1000}{\tau} = 0$$

which is a cubic of the form:

$$S^3 + aS^2 + bS + c = 0.$$

It is evident that the boundary condition between stable and unstable roots occurs when $ab = c$. For $ab > c$, the system oscillates with an exponentially decreasing amplitude and for $ab < c$, the system oscillates with an exponentially increasing amplitude.

Now, if the characteristic of the system can be altered such that $ab = c$, the correction will be a measure of the phase shift of the summing amplifier. Such a circuit is shown in Fig. 2. The transfer function for the correcting circuit (II) is:

$$e_0 = -\frac{10e_i}{S} - \frac{\alpha e_0}{S}$$

so that

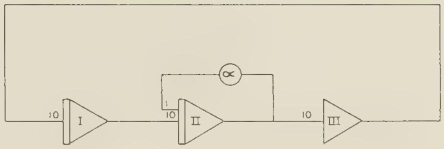


Fig. 2

$$\frac{e_0}{e_i} = -\frac{10}{\alpha + S}.$$

We then have

$$W(S) = \frac{-1000}{S(\alpha + S)(1 + \tau S)}$$

and the characteristic is:

$$S(\alpha + S)(1 + \tau S) + 1000 = 0.$$

Simplifying, we obtain

$$S^3 + \frac{\tau\alpha + 1}{\tau} S^2 + \frac{\alpha}{\tau} S + \frac{1000}{\tau} = 0.$$

The requirement that the coefficients satisfy the relation $ab = c$ gives us

$$\alpha = -\frac{1}{2\tau} \pm \sqrt{\frac{1}{4\tau^2} + 1000}.$$

For a typical amplifier with a gain of 1, τ is of the order of 8×10^{-6}.

Substituting this value of τ in the above equation we obtain a function of the following form:

$$-X \pm \sqrt{X^2 + \Delta} = f(x)$$

where Δ is a small increment. Now let

$$\sqrt{X^2 + \Delta} = X + \sigma.$$

Then

$$X^2 + \Delta = X^2 + 2X\sigma + \sigma^2.$$

To a first approximation σ^2 may be neglected and therefore $\sigma = \Delta/2X$.

Substituting back into the equation for α, we have for the positive root

$$\alpha \cong -\frac{1}{2\tau} + \frac{1}{2\tau} + 1000\tau \cong 1000\tau$$

and for small phase angles θ

$$\theta = \tan^{-1} \omega\tau \cong \omega\tau = \frac{\omega\alpha}{1000} \text{ radians.}$$

Since $\omega = 2\pi f$ and π radians equal 180 degrees, we obtain

$$\theta = \frac{9}{25} f\alpha \text{ degrees.}$$

This approximation is useful for system gains (K) up to 50×10^6, so that the general formula

$$\theta = \frac{\omega\alpha}{K} \text{ rad} \quad \text{or} \quad \theta = \frac{360 f\alpha}{K} \text{ degrees}$$

is valid within the range of K's encountered for which this method of measurement is useful.

Appendix

The following analysis will show that in the range of summing amplifier phase shift of 0.02° to 0.0003° and system frequencies of 2.5 to 90 radians per second, integrator error is negligible and the measured values of summing phase shifts are accurate within at least 1 per cent.

The circuit shown in Fig. 3 may be used to analyze both the integrator and summing amplifiers where A exceeds 50,000 for a good operational amplifier. If we apply the performance equation[1] for a parallel feedback amplifier, the transfer function F for the circuit of Fig. 3 is

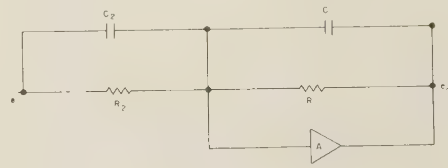


Fig. 3

$$F = \frac{R_1(1 + S\tau_2)}{R_2(1 + S\tau_1)}$$

where $\tau_2 = R_2C_2$ and $\tau_1 = R_1C_1$.

For a summing amplifier let $K_A = R_1/R_2$ where K_A is the amplifier gain factor. Then

$$F(j\omega) = \frac{K_A}{1 + \omega^2\tau_1^2}(1 + \omega^2\tau_1\tau_2 + j\omega(\tau_2 - \tau_1)).$$

The phase angle is given by

$$\tan\theta = \frac{\omega(\tau_2 - \tau_1)}{1 + \omega^2\tau_1\tau_2}$$

and since $\omega^2\tau_1\tau_2$ is small

$$\theta \cong \tan^{-1}\omega(\tau_2 - \tau_1) \cong \omega R_2(C_2 - K_AC_1).$$

For a gain of unity and a range of $C_2 - C_1$ from 2 to 8 $\mu\mu$f at $\omega = 31.6$ radians, θ will range from 0.0036 to 0.0145 degrees. Higher gains will increase the phase shift.

For an integrator $\tau_1 \gg \tau_2$. Typical values are $\tau_1 = 10^6$, $\iota_2 = 2 \times 10^{-6}$. The integrator phase shift error is given by

$$B = \tan^{-1}\frac{1 + \omega^2\tau_1\tau_2}{\omega(\tau_2 - \tau_1)} \cong -\left(\frac{1}{\omega\tau_1} + \omega\tau_2\right)$$

for $\tau_1 \gg \tau_2$.

The term $\omega\tau_2$ is insignificant with respect to summing phase shift since at $\omega = 31.6$ the contribution of this term is 0.000057° (for an integrator gain of 10). Higher gains will reduce this error proportionately. This compares with a typical summing phase shift of 0.007° (for a gain of 10). The gains and frequencies are those used in the system described in the body of this report.

The term $-(1/\omega\tau_1)$ becomes significant when ω approaches 2.23 radians at which time we have $\beta = \theta/10$ (assuming gains throughout the system approach 1).

H. Hamer and A. Lupinski
Electronic Associates, Inc.
Long Branch, N. J.

[1] Korn and Korn, "Electronic Analog Computers," McGraw-Hill Book Co., Inc., New York, N.Y. 2nd ed., sec. 4.2; 1956.

* Received by the PGEC, May 13, 1957.

High-Speed Digital Multiplication*

In a recent note[1] Estrin, Gilchrist, and Pomerene showed how carry-storage could be usefully combined with carry-propagation-detection circuits to produce very fast binary multipliers. The present note is intended to introduce a new method (tentatively named the "Modified Short-Cut" or, "MSC" multiplication process), which when combined with carry-propagation-detection circuits, produces equally fast or even faster multipliers. This new method is more completely developed, analyzed, and described elsewhere.[2,3]

The decimal short-cut multiplication method as used on desk computers is well known and its analog may be similarly used in binary arithmetic units. The process which, for example, expresses a sequence of "ones," $\sum_{r=m}^{n} 2^r$ as $(2^{n+1}-2^m)$, is however less efficient than a straightforward multiplier. Thus for an $(m+1)$-bit signed multiplier it requires, on the average, $(m+1)/2$ addition or subtraction operations (henceforth termed "operations") whereas the latter requires only $m/2$ additions. This inefficiency arises from multiplier-bit combinations of the form $\cdots 0100 \cdots$ and $\cdots 1011 \cdots$, that is, "isolated" bits. If the former combination occurs in a positive multiplier, for example, a straightforward multiplication requires only one addition of the multiplicand to the partial sum whereas a short-cut method performs a subtraction followed by a shift and an addition. Similarly when the second combination arises in a positive multiplier, a short-cut method demands two operations, whereas it is easily seen that a single subtraction would suffice. Similar inefficiencies arise when a negative multiplier represented in either "ones" complements or "complements to two" is controlling the multiplication. Recognition of these facts has led to the development of the MSC process in which a standard short-cut procedure (which "operates" whenever neighboring bits are unequal) is used except when "isolated-bits," appropriately defined, are encountered. When this occurs, then for positive or negative multipliers respectively, isolated ones lead to addition or subtraction and isolated zeros to subtraction or addition. A study of the process shows that it lends itself very easily to mechanization.

The method may be formalized by considering a ternary coding of binary numbers, such that

$$2^{-n}\sum_{0}^{m} b_t 2^t \equiv 2^{-n}\sum_{0}^{m+1} (-1)^{S_t} C_t 2^t \qquad (1)$$

where c_t and S_t are binary variables. *Any* multiplication method based on such a recoding will perform an operation in any cycle t in which the controlling bit is b_t and in which $(c_t=1)$. The sign of this operation is controlled by S_t, an addition being performed when $S_t=0$.

The MSC process may then be defined by expressions:

$$c_t \equiv (b_t \neq b_{t-1}) \,\&\, \bar{c}_{t-1} \qquad (2)$$

$$s_t \equiv (b_{t+1} \,\&\, b_t) v \; [(b_{t+1} \neq b_t) \,\&\, s_{t-1}] \qquad (3a)$$

$$\equiv (c_t \,\&\, b_{t+1}) v \; (\bar{c}_t \,\&\, s_{t-1}) \qquad (3b)$$

In these expressions b_t $(t<0)$ is implicitly defined in a manner which depends on the negative number representation used.

Eqs. (2) and (3a) follow immediately from the modification of the standard short-cut methods as outlined. Eq. (3b) may be obtained from an alternative consideration of the process or by an application of the appropriate algebra from (2) and (3a).

The value of S_t is clearly of no consequence when $c_t=0$, *i.e.*, when $\bar{c}_t=1$, and for a mechanization of the process a more convenient (though not equivalent) form of (3b) is

$$S_t = b_{t+1}. \qquad (4)$$

Tocher[4] has considered the most general relations governing ternary recodings of binary numbers of the type of (1) and shows by a proof a modified version of which is reproduced by Lehman[2] that the MSC process may be considered the optimum multiplication process (in a sense to be defined) in asynchronous parallel binary machines or in synchronous machines in which multiplication may be allowed to take a variable number of word (or bit) times. For such machines the optimum process is that which minimizes the number of operations for a given number length, that is, one that minimizes $\sum c_t$.

It is further shown by different methods by both Tocher[4] and Lehman[2] that the average number of operations when using the MSC process with an $m+1$ bit (signed) number is

$$\tfrac{1}{9}\{3m + 4 - [9 - (-1)^m](\tfrac{1}{2})^{m+1}\} \qquad (5)$$

and for the more usual values of m this approximates to $m/3$. Moreover, from (2) it follows immediately that $\overline{c_t \,\&\, c_{t-1}} \equiv 1$; that is, that operations can never be required in two successive cycles. Thus the *maximum* number of operations required is $[(m+2)/2]$ and this is approximately equal to the *average* number required in straightforward multiplication. In any cycle in which an operation has occurred, it is thus possible to conclude that cycle with a two-place shift back of the partial sum and the multiplicand, instead of the more usual single-place shift. This in turn implies that the actual average multiplication time (exclusive of store access) may be reduced by about 30 per cent. If such two-place shift-links are provided it is possible to further simplify the controlling network based on (2), to

$$c_t \equiv b_t \neq b_{t-1}. \qquad (6)$$

It is shown further[2,4] that the MSC process, as defined, is immediately applicable to signed multiplication irrespective of whether "ones" complements or "complements to two" are used for negative number representation. Naturally the addition circuits must be matched to this representation. This implies that when "ones" complements are used for negative number representation, provision must be made not only for end-around carry but if double-length or unrounded single-length products are required, either a double-length accumulator or some other simple circuit modification must be provided to prevent the possible introduction of an error whenever the operation and hence the partial product changes sign.

A detailed comparison of the speeds and hardware requirements of the various multiplication processes is difficult since the relative performances are very much dependent on circuit designs and circuit speeds. An attempt has, however, been made[2,3] to assess and compare the relative average speeds of the carry-storage system, the MSC system with two-place shift links, the MSC system with only single-place shift links, a straightforward multiplication method, and finally a combined MSC carry-storage multiplier. General expressions in terms of the response speeds of various parts of the circuits and actual times based on assumptions about these speeds are quoted for a variety of hypothetical machines which, in some cases, are assumed to include carry-detection circuits.[5]

In particular, one set of figures relates to the times quoted by Gilchrist, Pomerene, and Wong.[5]

And, the figure given for multiplication based on carry storage was originally 8.2 μsec. Further consideration[3] of Gilchrist's[5] figures has revised this estimate since the delay obtained from one carry network stage appears to be 0.120 μsec and not 0.045 μsec as originally quoted. In any case it is difficult to see how Estrin *et al.*[1] arrive at a figure of 6.4 μsec. Total average multiplication time in the unit described by them will surely be of order $41\times0.15+40\times0.12+0.21=11$ μsec and this is actually slower than the conventional multiplier.[5] By making the carry register a shifting register this average time may, of course, be reduced to $41\times0.15+20\times0.12+0.21=8.8$ μsec, since the addition of carries need then be performed only in those cycles in which the multiplier bit is significant. Comparable figures for the MSC process applied to a unit using carry-detection circuits are 7.0 μsec if additional two-place shift links are provided and 9.0 μsec if only the normal single-bit links exist.

This analysis is, however, only based on Gilchrist's figures and it might well be that with certain types of circuit elements the carry-storage process is faster and more appropriate. An example is a synchronous circuit based on (serial type) bit-synchronous circuit elements in which each such element involves a delay of an integral fraction (say $\frac{1}{2}$ or 1) of a bit time. This might occur for example in applications of magnetic core elements or where circuits designed for serial use are to be utilized in the construction of a parallel machine.

* Received by the PGEC, March 15, 1957; revised manuscript received, June 20, 1957.

[1] G. Estrin, B. Gilchrist, and J. H. Pomerene, "A note on high-speed digital multiplication," IRE Trans., vol. EC-5, p. 140; September, 1956.

[2] M. Lehman, "Parallel arithmetic units and their control," Ph.D. thesis, Univ. London, February, 1957.

[3] M. Lehman, "Short-cut multiplication and division in automatic binary digital computers with special references to a new multiplication process," submitted for publication to *Proc. IEE.*

[4] K. D. Tocher, "Technique of multiplication and division for automatic binary computers," (in preparation).

[5] B. Gilchrist, J. H. Pomerene, and S. Y. Wong, "Fast carry logic for digital computers," IRE Trans., vol. EC-4, pp. 133–136; December, 1955.

Finally, it should be added that the multiplication methods based on carry-storage and on the MSC process are not, of course, mutually exclusive. In a combined system it would clearly be desirable to make the carry-store a shifting register, so that the carries would be added in only in those cycles of multiplication in which an addition or subtraction (obtained by complementing the multiplicand register) of the multiplicand was also required. An estimate of the speed of the resultant multiplier, based on Gilchrist's figures[5] shows that times of 6.1 and 8.0 μsec would be obtained according as to whether the unit was or was not provided with two place shift links. The latter figure is greater than that obtained for the faster MSC unit; hence, it is clear that the combined system would only be used in a machine already equipped with two-place shift links. The resultant arithmetic unit would be some 25 per cent larger than one multiplying by the fast MSC process and would yield a 13 per cent increase of speed. This resultant 6.1 μsec average multiplication time is of the same order as the access time of presently envisaged high-speed (magnetic-core) stores; and it is therefore considered that with present day circuit techniques, this combined "Carry-Storage MSC Multiplier" with or without carry-detection is not worthwhile.

M. Lehman
Ministry of Defence
Tel Aviv, Israel
Formerly with Computer Lab.
Ferranti Ltd.
London, England

Contributors

D. D. Aufenkamp was born in Brock, Neb., on March 31, 1927. He received the B.A. degree from Southwestern at Memphis, Tenn., in 1949 and the degree of Doctor at d'Université from the University of Paris in 1952.

From 1946 to 1948, he served in the U. S. Navy as an aviation electronics technician. He was a member of the Ecole d'Eté de Physique Théorique, University of Grenoble, Les Houches, France, the summer of 1951.

Dr. Aufenkamp was instructor in mathematics at Princeton University from 1952 to 1953, and in physics at Southwestern the summer of 1953. He was an instructor in mathematics at Reed College from 1953 to 1956 and visiting assistant professor of electrical engineering at the University of Illinois from 1956 to 1957. He has served as a member of the technical staff of the Bell Telephone Laboratories, Inc. during the summers of 1954 and 1955. Dr. Aufenkamp was then a member of the applied mathematics staff of Systems Laboratories Corp., Sherman Oaks, Calif., for the summer of 1956, a part-time member of the staff during the subsequent academic year, and a permanent full-time member of the applied mathematics staff.

He is now associated with Lockheed Missile Systems Division, Sunnyvale, Calif.

❖

Martin R. Bates was born on November 23, 1925, in Brooklyn, N. Y. From 1928 to 1937, he was in Tientsin, China. After returning to the United States, he enlisted in the Navy V-12 program in 1943, and was sent to Cornell University, Ithaca, N. Y. He received the A.B. degree in mathematics and physics in February, 1946, and the M.S. degree in mathematics in 1949, both from Cornell University. Upon graduation in February, 1946, he was commissioned an Ensign in the U. S. Naval Reserve and completed a brief training cruise.

Mr. Bates was successively a teaching assistant, teaching fellow, and summer session instructor in mathematics at Cornell University from October, 1946, to August, 1949. He instructed in mathematics at Union College, Schenectady, N. Y., from September, 1949, to June, 1950, and then started work at Bell Aircraft Corporation, Buffalo, N. Y.

For the last five years, he has been investigating prediction problems in connection with an automatic landing system. From May, 1956, to February, 1957, he was the project engineer on a ship motion measurement program conducted on the *S. S. Mariposa*.

Mr. Bates is a member of the Mathematical Association of America.

❖

D. H. Bock was born on December 30, 1931, in Newark, N. J. From 1938 to 1947 he was in Dresden, Germany, and after his return to the United States, he obtained the B.A. degree in physics at the University of Buffalo, Buffalo, N. Y. in 1953.

At present, Mr. Bock is an electronics engineer at the Bell Aircraft Corporation, Buffalo, N. Y., where he is engaged in ship motion research and in the development of a ship motion predictor for an automatic carrier landing system.

❖

Frederick P. Brooks, Jr. (S'53–M'56) was born in Durham, N. C., on April 19, 1931. He received the A.B. degree in physics from Duke University, Durham, N. C., in 1953, and the S.M. and Ph.D. degrees in applied mathematics from Harvard University, Cambridge, Mass., in 1955 and 1956, respectively. His graduate work was in the design and application of computing systems, and his doctoral thesis dealt with the design of digital systems for business data processing.

Since 1956, Dr. Brooks has been associated with the IBM Corporation Product Development Laboratory in Poughkeepsie, N. Y., where he is at present a staff engineer in the Project 7000 planning group.

Dr. Brooks is a member of the Association for Computing Machinery, the Operations Research Society of America, the Institute of Management Sciences, Phi Beta Kappa, and Sigma Xi.

❖

R. A. Dandl was born in Cascade, Iowa, in 1922. He was associated with Oak Ridge National Laboratory, Oak Ridge, Tenn., from 1944 to 1951. In 1952, he received the B.S. degree in physics from Alabama Polytechnic Institute, Auburn, Ala. He then worked briefly for Melpar, Inc., and the Auburn Research Foundation before returning to Oak Ridge National Laboratory in 1954, where he has spent part time on the development of reactor control analogs. He is presently in the Thermonuclear Experimental Division.

❖

Lawrence J. Fogel (A'46–SM'53) was born in Brooklyn, N. Y., on March 2, 1928, He received the B.S.E.E. degree from New York University in June, 1948, did graduate study at the Polytechnic Institute of Brooklyn in 1948–49, and was graduated with the M.S. degree in electrical engineering from Rutgers University in June, 1952. He completed all courses for the Ph.D. degree in electrical engineering in June, 1954, at the Polytechnic Institute and took part in the 1954 summer communication theory program at Massachusetts Institute of Technology. He is working toward completion of a Ph.D. thesis in the field of biotechnology at the University of California at Los Angeles.

From June, 1948, to July, 1949, Mr. Fogel worked at the Watson Laboratories of the U. S. Air Force at Red Bank, N. J., and from July, 1949, until October, 1950, at the Armed Services Electro Standards Agency, Fort Monmouth, N.J. He joined the Coles Signal Laboratory, U. S. Army Signal Corps, Fort Monmouth, and was there until June, 1953, when he became affiliated with Davenco Inc., New York, where he did highly

classified government research until April, 1954. He was then a senior research engineer with Stavid Engineering Inc., Plainfield, N. J., from April, 1954, until June, 1956.

Mr. Fogel is now associated with Convair where he has been assigned the responsibility for human engineering systems analysis and design for manned aircraft. At Convair, he has devised a radical new display and control cockpit which for the first time offers intersensory compatibility as protection against vertigo onset.

Mr. Fogel is a member of the Acoustical Society of America and the American Rocket Society.

❖

Franz E. Hohn (A'54) was born in Warrenton, Mo., on September 4, 1915. He received the B.S. degree in mathematics from McKendree College, Lebanon, Ill., in 1936 and the M.S. and Ph.D. degrees in mathematics from the University of Illinois, Urbana, Ill., in 1937 and 1940.

He has taught at the University of Arizona, Tucson; Guilford College, N. C.; the University of Maine, Orono, and at the University of Illinois where he is currently associate professor of mathematics.

He has been a research guest and consultant at Bell Telephone Laboratories. His writings are in the fields of matrix algebra and switching theory.

❖

Albert L. Hopkins, Jr., was born on May 6, 1931, in Chicago, Ill. He attended Harvard University, Cambridge, Mass., where he received the A.B. and A.M. degrees in applied mathematics, and the Ph.D. degree in applied mathematics in 1953, 1954, and 1957, respectively. His graduate work was in the design and application of computing systems, and his doctoral thesis dealt with investigations of computing machine components. He is presently instructor in applied mathematics at Harvard, teaching courses in control systems engineering.

❖

Warren S. McCulloch (SM'54) was born in Orange, N. J., on November 16, 1898. He received the B.A. degree from Yale University in 1921, and the M.A. degree from Columbia University in 1923, later studying medicine at Columbia, where he was awarded the M.D. degree in 1927.

During 1927–1928, Dr. McCulloch was an interne and resident at Bellevue Hospital's Second Division, Neurological Service. From 1928 to 1930, he worked with experimental epilepsy at the Neurosurgical Laboratory and Department of Neurology, Columbia. His graduate work in mathematical physics at New York University in 1931–1432 was interrupted by illness.

From 1932 to 1934, Dr. McCulloch worked in the admission service at Rockland State Hospital, Orangeburg, N. Y. He attended Yale University's Laboratory of Neurophysiology in 1934. He was an Honorary Research Fellow in 1934–35, a Sterling Fellow in 1935–36, an instructor in 1936–40, and an assistant professor as a research assistant in 1940 41.

Dr. McCulloch joined the Department of Psychiatry at the University of Illinois College of Medicine in 1941, where he became director of the Laboratory for Basic Research in Psychiatry. He held several posts there, finishing his tenure as a professor of psychiatry and physiology and as a neurophysiologist. He left the University of Illinois to accept his present position as a staff member of the Research Laboratory of Electronics, Massachusetts Institute of Technology, Dr. McCulloch has written on such subjects as functional organization of the brain and facilitation, extinction, and functional organization of cerebral cortex.

❖

George A. Miller was born in Charleston, W. Va., on February 3, 1920. He received the B.A. and M.A. degrees from the University of Alabama, Tuscaloosa, in 1940 and 1941, respectively. He then received the A.M. and Ph.D. degrees from Harvard University, Cambridge, Mass., in 1944 and 1946, respectively.

Dr. Miller was a research fellow at Harvard from 1944 to 1948, and served as assistant professor of psychology there from 1948 to 1951. From 1951 to 1955 he was at the Massachusetts Institute of Technology as an associate professor, returning then to Harvard.

Dr. Miller has been published in the fields of audition, perception of speech, psychology of communication, and is the author of the book, *Language and Communication*.

He is a member of the American Academy of Arts and Sciences, American Association for the Advancement of Science, American Psychological Association, American Statistical Association, Linguistic Association of America, Psychometric Association, Society of Experimental Psychologists, Acoustical Society of America, American Physiological Society, and Eastern Psychological Association.

❖

Peter G. Neumann (S'56) was born on September 21, 1932, in New York, N. Y. He received the A.B. degree in mathematics in 1954 and the S.M. degree in applied mathematics in 1955, both from Harvard University, Cambridge, Mass. He is currently a Ph.D. candidate, doing research in switching theory, and is a teaching fellow and research assistant at the Harvard Computation Laboratory. His special interests include design and organization of computers and problems of coding and communication in information theory.

Mr. Neumann is a member of the Association for Computing Machinery.

❖

Lazarus G. Polimerou (S'52–A'53) was born in Glen Falls, N. Y., on January 28, 1929. He received the B.S.E.E. degree from Northeastern University, Boston, Mass., in 1952, and is currently completing requirements for the M.S.E.E. degree from Southern Methodist University, Dallas, Texas.

During 1951 to 1952, he was employed by the U.S. Navy Underwater Sound Laboratory. From 1952 to 1953 he worked with the U. S. Army Ordnance, EMLD Laboratories, at White Sands Proving Ground, where he was instrumental in setting up the first analog computer installation and later took charge of the Analog Group.

He entered the U.S. Army for military duty in 1953, serving as an instructor in electronics at Fort Monmouth, N. J. In 1954, he returned to the Analog Computing Laboratory at White Sands. After completing his tour of duty in 1955, he joined Convair, Fort Worth, Texas, engaging in the design and application of computer techniques in the solution of analog problems.

Mr. Polimerou is a member of the AIEE.

❖

Fred Powell (SM'57) was born in Boston, Mass., on June 14, 1922. He received the B.A. degree in mathematics from Harvard University, Cambridge, Mass., in 1944.

In 1947, after three years of war service in Naval aerology and seismological research, he joined the Bell Aircraft Corporation, Buffalo, N. Y., where he has been engaged in system dynamic analysis and analog computation. Since 1951, his work has been concentrated on automatic control system theory and related fields.

Mr. Powell is a member of the Institute of Aeronautical Sciences.

❖

Henry Quastler was born in Vienna, Austria, on May 21, 1908. He attended the University of Vienna and was graduated with the M.D. degree in 1932. Dr. Quastler was appointed radiologist at the General Hospital in Tirana, Albania, and held that post for five years until he came to the United States in 1939.

Dr. Quastler served as assistant radiologist at New Rochelle Hospital from 1940 to 1942, and assumed the post of radiologist at the Carle Hospital Clinic in Urbana, Ill., until 1949. He became associate professor of radiobiology at the University of Illinois in 1949, and later an associate professor of physiology. In 1955, Dr. Quastler was named a research professor in the Control Systems Laboratory, and the same year became senior radiobiologist at the Argonne National Laboratory.

He assumed his present position as senior radiobiologist at the Brookhaven National Laboratory, Upton, N. Y., in 1956. His main interests are theory of organization and the effects of radiation on living creatures.

❖

J. H. Sasseen (M'53) was born November 21, 1921, in Texas. He received the B.S. degree in electrical engineering from the Rice Institute in February, 1944. From 1944 to 1946, he served in the U. S. Navy as a radar and radio officer in the Pacific area. In 1946, he was employed by the Humble Oil & Refining Company to work with radar, radio communication equipment, and seismic exploration equipment in the Gulf of Mexico. After four years, he was assigned to Hum

ble's Geophysical Engineering Group where he worked with the design and construction of the electronics for seismic and well-logging equipment. During the last few years, he has been assigned to Geophysics Research where he has worked on the project of designing, constructing, and developing a dip-log cross correlator.

Mr. Sasseen is a member of the Society of Exploration Geophysicists, and is a registered professional engineer in the State of Texas.

❖

George S. Sebestyen (M'55) was born on November 27, 1931, in Budapest, Hungary. He received the B.S. and M.S. degrees in electrical engineering at the Massachusetts Institute of Technology in 1955. Under the cooperative plan in electrical engineering, during the years 1951 to 1954, he worked at the General Radio Company. In his graduate year, he was a teaching assistant, and later, a research assistant at M.I.T.

Upon graduation, Mr. Sebestyen joined the Research Department of Melpar, Inc., Boston, Mass., where he is engaged in research concerned with the military applications of the newer concepts in communication theory.

He is a member of Eta Kappa Nu and Sigma Xi.

❖

Sundaram Seshu (A'54) was born in Madras, India, on June 1, 1926. After completing his undergraduate work in physics and electrical engineering at the University of Madras and the Indian Institute of Science, respectively, he joined the Graduate College of the University of Illinois, Urbana, Ill., in 1952.

In 1955, he obtained the degree of Ph.D. in electrical engineering and served on the faculty of the University of Illinois until September, 1956. Since that time, he has been on the faculty of Syracuse University, Syracuse, N. Y.

❖

Robert S. Stone (A'55) was born in Fond du Lac, Wis., on February 16, 1923. He received the B.S. degree in physics from the California Institute of Technology, Pasadena, in 1948. During 1943–1946, he served in the U.S. Army (ASTP and Signal Corps), and since 1948, he has been at the Oak Ridge National Laboratory, Oak Ridge, Tenn., working on various aspects of the measurement of radiation. Since 1954, he has been associated with the analog facility of the Reactor Controls Department of that Laboratory, where he has been concerned with simulator studies of problems in reactor operation and control.

Mr. Stone is a member of the Southeastern Simulation Council.

❖

Thomas B. Van Horne (S'51–A'52) was born March 24, 1929, in Buenos Aires, Argentina. He received the B.S. degree in physics in 1951 from the University of Minnesota, Minneapolis, Minn., and joined the Guided Missile Laboratory of the Hughes Aircraft Company, Culver City, Calif., that same year, where he worked on the Falcon Missile Project. In 1955, while still at Hughes, he received the M.S. degree in applied physics from the University of California at Los Angeles.

Since then he has been a member of the technical staff of the Systems Analysis Laboratory at Hughes, where he is presently engaged in effectiveness studies on airborne fire control systems.

❖

William V. Wright (S'52) was born in Winston-Salem, N. C., on September 15, 1931. He received the B.S.E.E. degree from Duke University, Durham, N. C., in 1953 and the S.M. degree from Harvard University, Cambridge, Mass., in 1954. At present, he is completing the requirements for the Ph.D. degree in applied mathematics at Harvard University.

During the summers of 1952 through 1954, he was employed by the Westinghouse Electric Corporation and the Bell Telephone Laboratories, where he worked in the development of pulse circuitry, and at North American Aviation, Inc., where he did systems analysis.

He was granted a Fulbright scholarship for study at the Mathematisch Centrum in Amsterdam, the Netherlands, during 1956 and 1957. While there, he participated in the logical design of a large-scale electronic computer, the XI.

Mr. Wright is a member of Phi Beta Kappa, Tau Beta Pi, and Eta Kappa Nu.

PGEC News

Joint Computer Conference Proceedings Available

A few copies of the Proceedings of past Joint Computer Conferences are still available from the IRE Headquarters, 1 East 79th Street, New York 21, N. Y. A list of available Proceedings and their prices appears below. Orders should be accompanied by remittances payable to the Institute of Radio Engineers.

1952 Review of Electronic Digital Computers (covering the December, 1952 Conference) $3.50
1954 Proceedings of the Eastern Joint Computer Conference $3.00
1955 Proceedings of the Eastern Joint Computer Conference $3.00
1956 Proceedings of the Eastern Joint Computer Conference $3.00

Fellowship Sponsored by PGEC

The National Academy of Sciences-National Research Council has announced the awarding of a predoctoral fellowship to Eugene Walter Henry, presently with the United States Air Force at Dayton, Ohio.

This fellowship is sponsored by the Professional Group on Electronic Computers of the Institute of Radio Engineers and is administered by the NAS-NRC. It is awarded to an outstanding young scientist to continue graduate training in electronic computers.

Mr. Henry was born in Wichita Falls, Texas, in 1932. He received the B.S.E.E. degree *magna cum laude* from the University of Notre Dame in 1954, and the M.S.E.E. degree in 1955 from the same university. He majored in electrical engineering with a minor in electronics. Since graduation he has been on active duty in the Air Force as a lieutenant, assigned to the Communication and Navigation Laboratory, Wright Air Development Center, Dayton, Ohio. His tour of duty ended in June, 1957.

He plans to pursue his fellowship in the Department of Electrical Engineering at Stanford University (California), beginning in September, 1957. His present program of study involves advanced training and research in servomechanism electronics and will include electronic computer theory and advanced network theory to facilitate servosystem solutions. He will be working toward the Ph.D. degree, but his long-range goal is "to build, with continued study and research in industry or universities, systems with superior operational and environmental characteristics."

Chapter News

Washington, D. C.

The following officers have been elected for 1957–1958:

Chairman, R. G. Lilly, Bureau of Ships, Navy Department
Vice-Chairman, L. D. Whitelock, Bureau of Ships, Navy Department
Secretary, B. J. Ellis, Department of Defense.

Present plans call for five meetings in 1957–1958, to be held on the first Wednesday of October, November, February, March, and May.

Chicago

This chapter has no secretary, and the name of H. E. Kantner was listed in error in the March, 1957 issue.

Computer Symposium

The AIEE Feedback Control Sytems Committee with participation by the IRE Professional Group on Automatic Control and the ASME-IRD is organizing a symposium on the theme of "Computers in Control" at the Chalfonte-Haddon Hall Hotel in Atlantic City, N. J., on October 16–18, 1957. The symposium will emphasize the use of digital and analog computers both as elements of feedback control systems and as utilized in the design of such systems.

Some 35 technical papers by leading scientists in this country and by Prof. J. Tsypkin of The Academy of Science, Moscow, U.S.S.R., Dr. M. Pélégrin of The Ministry of Defense, France, and Dr. Blackman of The Imperial College of Science and Technology, London, England will be prepresented.

A portion of the papers to be presented at the conference are available in preprint form from the AIEE; the *Proceedings* of the Conference will be published by the AIEE early in 1958. Conference arrangements have been made by a committee consisting of Harold Chestnut of GE, E. W. Grabbe, Ramo-Wooldridge, Adam Kegel, Westinghouse, John Ragazzini, Columbia University, and E. Mishkin and Chairman J. G. Truxal, both of Polytechnic Institute of Brooklyn.

PGEC Fellowship

For the second time, the PGEC will sponsor an annual Fellowship for graduate study in computing. The amount of the stipend will be "2,000.00, plus tuition not to exceed "1,000.00.

This Fellowship will be administered through the Office of Scientific Personnel of the National Academy of Sciences, and the selection of the successful recipient will be made by that office. There will be no restrictions as to the university or college to be attended, but it will be necessary for the Fellow to specialize in computing or a closely allied field.

Applications for this Fellowship will be received through January, 1958, and should be addressed to:

Fellowship Office
Office of Scientific Personnel
National Academy of Sciences
2101 Constitution Avenue, N.W.
Washington 25, D. C.

Announcements made by the National Academy of Sciences will not mention this Fellowship by name but will merely indicate that the National Academy Fellowship Program has available a number of fellowships of which this is one. The PGEC is publicizing this Fellowship directly through its own Transactions, the IRE Student Quarterly, Proceedings of the IRE, and through each PGEC Chapter.

Richard W. Melville, *Chairman*
Student Activity

Reviews of Current Literature

Copies of books or of articles to be reviewed should be sent to H. D. Huskey, Department of Electrical Engineering, University of California, Berkeley, California. The editors wish to express their gratitude to the reviewers who, through their efforts, make this section possible.—*H. D. Huskey.*

GENERAL

57-85

The Astonishing Computers—William B. Harris. (*Fortune*, pp. 136–139, 292–298; June, 1957.) The computer business is analyzed by the author with a description of digital computer manufacturers and the diverse use of their machines; simulation of jet fleet operations, scheduling of product shipping, predicting labor, material and parts requirements, optimizing oil refinery operation, and others. The problems that these computer manufacturers have and what they are doing about them, for example, how each of the large manufacturers is handling such problems as selling, financing, obsolescence, and manufacturing, are discussed along with current and future markets for computers.

J. A. Fingerett

57-86

A Survey of Domestic Electronic Digital Computing Systems, PB 111996—Martin H. Weik. (Reprint of Ballistic Res. Lab. Rep. No. 971, U. S. Dept. of Commerce, Office of Tech. Services, Washington, D. C., vii+272 pp.; 1956.) It is certainly not easy to remain abreast of developments in the electronic computer field. This survey helps to bring up-to-date (end of 1955) a broad picture of the field in the United States. Previous surveys of a similar nature date back to 1953 and are now obsolescent. Eighty-four "Domestic Electronic Digital Computing Systems" are described in the report. Two foreign manufacturers, Ferranti and Olivetti, are represented, presumably because these companies have U. S. sales organizations. On the other hand, at least some U. S. computers are missing, for example Harvard Mark IV, UNIVAC 120, and UNIVAC File Computer. It would be highly desirable if all foreign computers could be added to the list together with annual supplements. The descriptive information on each system runs from 2 to 4 pages and generally includes photographs, manufacturer, user(s), arithmetic and logical organization, size, cooling and power, input-output and memory data, production record and price, and operating experience. Much of the information was obtained from manufacturers and users. For this reason, as the editor points out, care must be taken when comparing detailed minute points since the contributers may have used different frames of reference. Anyone considering the acquisition of a computer could use the information in the survey to select a small group of computers for further consideration. The survey would be very useful for familiarization purposes. A chapter on Analysis and Trends follows the system description. The machines are listed in a number of ordered tables according to the characteristics of word length, add time, memory capacity, memory access time, tube and diode quantities, power requirements and cost. In addition the editor gives his opinion of trends in the field. A brief bibliography and a glossary of computer terms are included. (See 57-87, this issue.)

Douglas L. Hogan,
Courtesy of *Mathematical Tables and other Aids to Computation*

57-87

A Second Survey of Domestic Electronic Digital Computing Systems—Martin H. Weik. (Ballistic Res. Lab. Rep. No. 1010, Ballistic Res. Labs. Aberdeen Proving Ground, Md., 453 pp.; June, 1957.) One hundred and three computer systems are described in this report. Each system is described with respect to some sixteen headings. These headings were chosen to provide a measure of quality of the various systems and to indicate the position for those machines which are commercially available. The information is compiled from data supplied by manufacturers and, as the author is quick to admit, it is somewhat difficult to be sure that one has a valid picture in all cases. Various tables are given in an attempt to compare the characteristics of the computer systems. The tables include a list of manufacturers, number of systems produced, word length, operation times, memory access times and capacity, tube, crystal and transistor quantities, power requirements, and approximate cost. The power of the various computer systems is poorly indicated by such figures as memory access time, or arithmetic times. The reviewer would prefer to see the times required to solve by elimination, say, a system of linear equations of a given order both using fixed point arithmetic and floating-point (programmed) arithmetic. The number of systems using various methods of checking are summarized, and somewhat more academic figures are given such as the number of tubes per kilowatt and the price per tube for systems. This report is the best source for acquiring a picture of the present state of automatic computers in U.S.A. This report supersedes BRL Report No. 971 (see 57-86, this issue).

H. D. Huskey

621-32:621.9 57-88

Machine Tool Control—C. K. Marklew. (*Elect. Rev., Lond.*, vol. 159, pp. 189–193; August 3, 1956.) This is a brief survey of electronic control systems using punched tape, cinematograph film, magnetic tape, photocell devices, etc.

Courtesy of PROC. IRE
and *Wireless Engineer*

57-89

La Tecnica Britannica nel Campo delle Calcolatrici Elettroniche Automatiche Alla Luce del Recente Congresso di Londra—Paolo Ercoli, Giorgio Sacerdoti, and Roberto Vacca. (*Ricerca Sci.*, vol. 26, pp. 2321–2339; 1956.) This is an account of a series of visits to the computing laboratories of England with very brief descriptions given of a half-dozen large scale computers.

D. H. Lehmer
Courtesy of *Mathematical Reviews*

57-90

Analog Versus Digital Techniques for Engineering Design Problems—D. B. Breedon. (IRE TRANS., vol. IE-4, pp. 86–89; March, 1957.) Nearly every problem encountered in engineering at some time proceeds from the qualitative to the quantitative phase where the results of mathematical analysis must be applied in actual computation. Most often the computation is short enough that automatic means are not necessary. However, more and more problems are requiring powerful aids to calculation. This increase is due as much to expanded thinking encouraged by the mere availability of computers as to any actual backlog of work. Therefore it is to the engineer's advantage to know what computers can do for him, even though he may take his problem to someone else for final preparation and programming. The following text presents some examples in which automatic calculation is being used. The logic used in choosing the computing methods is shown based on the characteristics of problem and computer. As background for the examples the most important of these characteristics is presented briefly in the next section.

Courtesy of PROC. IRE

681.142 57-91

Industrial Data-Reduction and Analogue-Digital Conversion Equipment—P. Partos. (*J. Brit. Inst. Radio Eng.*, vol. 16, pp. 651–678; December, 1956.) A review of typical system specifications is presented and several existing and proposed installations are described.

Courtesy of PROC. IRE
and *Wireless Engineer*

This page has been left blank in order that readers may mount all reviews on cards.
—*The Editor*

621-52:681.142 57-92

Ship Stabilization: Automatic Controls, Computed and in Practice—J. Bell. (*Proc. IEE*, Part B, vol. 104, pp. 20–26; January, 1957.) Predictions by step-by-step and analog methods are given, the functioning of the analog computer being described, and examples of results presented. Some practical results from sea experience and a brief account of a stabilizing demonstration on a model are included.

Courtesy of Proc. IRE and *Wireless Engineer*

ANALOG COMPONENT RESEARCH

57-93

A Multipurpose Electronic Switch for Analog Computer Simulation and Autocorrelation Applications—N. D. Diamantides. (IRE Trans., vol. EC-5, pp. 197–202; December, 1956.) A system of four diodes in a series-parallel connection is combined with dc operational amplifiers in order to accomplish a variety of computational operations. The diode circuit is equivalent to a SPST switch survey or a voltage pulse. When inserted in series with the input of an amplifier or an analog memory, the switch makes possible waveform sampling or waveform quantizing of the input voltage. Other functions, such as the fast discharge of an integrator, are also achieved. A very significant application of the diode switch in computer circuitry is its use in combination with a multiplier and a storer (a bank of integrators) in order to obtain autocorrelation or cross-correlation of messages after they have been translated into voltages. A commutator or a ring counter is employed to provide the switching pulse. The correlator has the advantage of generating the correlation function concurrently with the message without necessitating previous recording and repeated playback.

Courtesy of Proc. IRE

631.142:621.383 57-94

Photoformer Analysis and Design—E. Elgeskog. (*Chalmers tek. Högsk. Handl.*, no. 172, p. 40; 1956.) Analysis is presented permitting the design of a photoformer with a bandwidth of several kc, suitable for use in a high-speed repetitive electronic differential analyzer. Difficulties due to the short response time are discussed in detail. Results obtained with an experimental system using a plane cr tube screen and photocathode are in good agreement with the theory.

Courtesy of Proc. IRE and *Wireless Engineer*

ANALOG EQUIPMENT

681.142 57-95

The Design and Applications of a General-Purpose Analogue Computer—R. J. A. Paul and E. L. Thomas. (*J. Brit. IRE*, vol. 17, pp. 49–73; January, 1957.) The design and construction are considered in detail with particular emphasis on amplifier gain, bandwidth, and phase shift. A computer whose design is based on this analysis and has some novel features is described. Its use on a variety of problems illustrates its wide range of application. (See 56-169, December, 1956 issue.)

Courtesy of Proc. IRE and *Wireless Engineer*

681.142 57-96

Radioactive-Fall-Out Computer—(*Electronic Ind. and Tele-Tech.*, vol. 15, pp. 57, 136; September, 1956.) Electronic analog techniques are used in a new computer to make rapid predictions of fall-out based on wind velocities and particle sizes. (See 57-7, March, 1957.)

Courtesy of Proc. IRE and *Wireless Engineer*

681.142 57-97

A New Computing Method using High-Frequency Currents—H. J. Uffler. (*Ann. Radioélect.*, vol. 11, pp. 187–199; July, 1956.) An electromechanical process is described for performing algebraic operations in analog computers. The system operates at 472 kc and comprises only passive components. Considerable accuracy and stability are achieved.

Courtesy of Proc. IRE and *Wireless Engineer*

UTILIZATION OF ANALOG EQUIPMENT

681.142 57-98

Applications of a Transformer Analogue Computer—J. R. Barker. (*Brit. J. Appl. Phys.*, vol. 7, pp. 303–307; August, 1956.) Use of the Blackburn analyzer for extracting latent roots of matrices, locating zeros of polynomials, and solving linear and nonlinear simultaneous equations is discussed.

Courtesy of Proc. IRE and *Wireless Engineer*

57-99

Ideal Transformers in the Synthesis of Analog Computer Circuits—R. H. MacNeal and G. D. McCann. (*Proc. 1955 Western Joint Computer Conf.*, pp. 16–23; 1955.) This article describes the use of transformers in electric analog computers to solve sets of algebraic equations having constant, real coefficients. The authors indicate that very nearly ideal transformers can be constructed and that, therefore, the techniques discussed in the article have practical significance. They show, using Cauer networks, that any coefficient matrix that can be transformed into a diagonal matrix with a triangular transformation matrix can be synthesized with the use of ideal transformers. This condition is much less restrictive than the conditions for realizability of a transformerless resistance-analog network and, therefore, more types of sets of equations can be solved. Also, with the use of transformers coordinate transformations can be performed. These cannot, in general, be performed with transformerless resistance networks. In many cases, physical measurements can be made on a mechanical system only if the system is restrained from rigid body movement. In these cases, electrical analogs must be synthesized using the information obtainable from the restrained physical system. This can be done using a technique described in this paper. Finally, some brief examples of the application of these techniques are described.

Peter G. Pantazelos

57-100

An Error Analysis of Electronic Analog Computers—V. A. Marsocci. (IRE Trans., vol. EC-5, pp. 207–212; December, 1956.) Due to the physical unrealizability of electronic adding and integrating circuits with ideal characteristics, errors will be introduced in the solution of differential equations obtained by the use of electrical analog computers. Numerical errors in the solution will be introduced by fluctuations in the value of plate and of grid supply voltages, changes in the values of circuit components, and changes in the values of the vacuum tube constants. In addition, the limited frequency response of the machine components will cause the computer to solve a characteristic equation of a higher order than the original characteristic equation whose solution is desired. The error in the solution manifests itself as a shifting in the roots of the original characteristic equation as well as the production of some extra roots. The effect of this change in the root position as well as the presence of the extra roots is experienced in the curve of the solution as a function of the independent variable. In a paper on the accuracy of differential analyzers, Macnee has derived an expression which gives the value of the characteristic root shift. The use of this expression is accurate only for certain types of ordinary differential equations. In this paper a new expression for the value of the root shift is derived. The analysis preceding the new root-shift expression is developed in such a manner as to include the Macnee analysis as a special case.

Courtesy of Proc. IRE

681.142:512.831 57-101

Analog Computer Synthesis and Error Matrices—P. M. Honnell and R. E. Horn. (*Commun. and Electronics*, pp. 26–32; March, 1956.) The performance of analog networks is studied by means of matrix analysis; errors and stability are evaluated.

Courtesy of Proc. IRE and *Wireless Engineer*

581.142:512 57-102

Solution of Algebraic Equations on an Analogue Computer—C. R. Cahn. (*Rev. Sci. Instr.*, vol. 27, pp. 856–858; October, 1956.) The roots of the equation are found with an electronic differential analyzer, using a method in which the operator adjusts a gauged potentiometer while observing a transient on a voltmeter or oscilloscope. Factors affecting the accuracy are discussed.

Courtesy of Proc. IRE and *Wireless Engineer*

57-103

The Solution of Algebraic, Transcendental and Integral Equations by Means of Analog Computers—I. M. Vitenberg and E. A. Gluzberg. (*Avtomat. i Telemeh.*, vol. 17, pp. 590–600, appendix to no. 7, 2.; 1956. Russian. English summary.) The authors

This page has been left blank in order that readers may mount all reviews on cards.
—*The Editor*

propose to solve the equations by minimizing the sum of absolute values of residuals, and for the minimization suggest auxiliary devices for which circuit diagrams are given.

A. S. Householder
Courtesy of *Mathematical Reviews*

57-104

Representation of Nonlinear Functions by Means of Operational Amplifiers—Robert M. Howe. (IRE TRANS., vol. EC-5, pp. 203–206; December, 1956.) The representation of a wide variety of nonlinear functions by means of the interconnection of unstabilized operational amplifiers is discussed. The nonlinear functions described include rectification, saturation functions, coulomb friction, dead space, and starting friction. The use of operational amplifiers alone to produce square waves and triangular waves, as well as gating operations, is also discussed. These latter circuits are combined to give a time division multiplier using only standard operational amplifiers as components. Accuracy capabilities for all of these nonlinear operations are the order of 0.01 to 1 per cent.

Courtesy of PROC. IRE

57-105

Representation of Functions of Several Variables by Means of Adders, Multipliers and Simple Functional Instruments—I. Š. Pinsker. (*Trudy Inst. Masinoved. Sem. Točn. Mašinostro. Priborostr.*, vol. 8, pp. 35–51; 1955. Russian.) Functions of one or two independent variables can be simply and precisely mechanized. Process of addition and multiplication may also be simply mechanized. Author proposes a scheme for simple mechanization of a function of n variables by approximating it with sums and products of functions which individually contain not more than two of the n variables. Thus, $f(x, y, t)$ may be represented by sums of terms of the following types:

$$\alpha(x)\cdot\beta(y)\cdot\tau(t);\ \alpha(x, t)\cdot\beta(y, t)\cdot\tau(x, y);\ \alpha(x)\cdot\beta(y, t);$$

$\alpha(x, y)\cdot\beta(x, t)$; etc. Only enough terms are selected to give the desired precision, and this is done by successive compensation to minimize the number of terms. A numerical example converts a set of tabulated values for $f(x, y, t)$ to the form $\beta_1(y, t)+\alpha_2(x)\cdot\beta_2(y, t)$ with 5-place precision.

W. W. Soroka
Courtesy of *Mathematical Reviews*

DIGITAL COMPONENT RESEARCH

57-106

Junction Transistor Switching Circuits for High-Speed Digital Computer Applications—G. J. Prom and R. L. Crosby. (IRE TRANS., vol. EC-5, pp. 192–196; December, 1956.) This paper describes junction transistor switching circuits capable of reliable operation at a clock rate of one megacycle. These circuits, consisting of a flip-flop, a gated pulse amplifier, and diode gates, consume a minimum of power and operate over a temperature range of $-55°C$ to $+85°C$ with complete transistor interchangeability. Applications of these circuits to binary counters, shift registers, and accumulators are also presented.

Courtesy of PROC. IRE

621.318.57:621.314.7 57-107

An Asymmetrical Bistable Circuit using Junction Transistors—(*Mullard Tech. Commun.*, vol. 2, pp. 254–278; July, 1956.) Conditions for the stable states of the basic switching circuit are analyzed and an empirical method for investigating the dynamic operation is presented. Detailed procedure for the design of a particular modified circuit is indicated. Reliable switching times of the order of 4 μsec may be obtained with repetition rates up to a fifth of the grounded-base cutoff frequency; trigger sensitivity is good.

Courtesy of PROC. IRE
and *Wireless Engineer*

621.375.4.018.7:621.314.7:681.142 57-108

Transistor Pulse Regenerative Amplifiers—F. H. Tendick, Jr. (*Bell Sys. Tech. J.*, vol. 35, pp. 1085–1114; September, 1956.) Design techniques are presented for synchronized regenerative amplifiers operating at a pulse repetition rate of the order of 1 mc and suitable for use in digital computers.

Courtesy of PROC. IRE
and *Wireless Engineer*

621.394:621.376.56:621.375.4.018.7:314.7 57-109

Transistorized Binary Pulse Regenerator—L. R. Wrathall. (*Bell Sys. Tech. J.*, vol. 35, pp. 1059–1084; September, 1956.) A simple repeater circuit is described which is suitable for use in a 12-channel pcm system over substantial lengths of transmission line. The system is arranged so that distortion due to LF cutoff in the output of one repeater is compensated in the next repeater, special feedback connections being provided for this purpose. Some performance figures and oscillograms are presented. The effect of interference on the production of errors is discussed.

Courtesy of PROC. IRE
and *Wireless Engineer*

57-110

The Effect of Collector Capacity on the Transient Response of Junction Transistors—J. W. Easley. (IRE TRANS., vol. ED-4, pp. 6–14; January, 1957.) The effect of collector depletion layer capacity on the transient response of junction transistors to a current input is calculated for the case of a resistive load. Expressions are given for the small-signal rise time of the common base, emitter, and collector configurations, and for the large-signal turn-on and decay times of the common-emitter and collector configurations. The analysis shows that the transient durations under most conditions of operation are approximately $(1+\omega_\alpha R_L C_c)$ times those which would be predicted for the short-circuit output approximation reported by Moll. Experimental results are reported which exhibit an excellent agreement with the analysis over a wide range of $\omega_\alpha R_L C_c$. An empirical examination has been made of the dependence of large-signal switching time on the range of operation point excursion. A satisfactory approximate representation of this dependence is provided by a first-order correction factor, which takes into account the functional dependence of ω_α and C_c on collector voltage.

Courtesy of PROC. IRE

57-111

High-Speed Gating Circuit Using the E80T Beam Deflection Tube—L. Sperling and R. W. Tackett. (IRE TRANS., vol. ED-4, pp. 59–63; January, 1957.) This paper describes a high-speed gate circuit for an information sampling system employing the E80T beam deflection tube. The time required to open this gate fully is less than 7 mμsec. This gate is superior to conventional multigrid gates in that the circuits are less complex, more reliable, and have only minimal signal feed-through. Furthermore, a preliminary stage of pulse stretching is available without additional circuit elements.

Courtesy of PROC. IRE

57-112

Pulse Generator and High-Speed Memory Circuit—Z. Bay and N. T. Grisamore. (IRE TRANS., vol. EC-5, pp. 213–218; December, 1956.) Circuits for the recycling of pulses by means of a driving tube and an electromagnetic delay line have been developed. The necessary characteristic for the driving tube is shown and the effects of the delay line on the amplitude and width of pulses with respect to recycling operation are explained. Two modes of operation of these circuits are possible. One mode of operation allows any number of pulses in a recycling period, the number being limited only by the time space on the delay line. The other mode restricts the number of pulses in a recycling period to a particular value. Experimental circuits are shown which have been used as generators of pulses as short as 5 mμsec at frequencies as high as 50 mc. Other circuits are shown which can be used as memory circuits for the storage of a number of these short pulses.

Courtesy of PROC. IRE

621.318.57:621.387 57-113

The Design of Cold-Cathode-Valve Circuits—J. E. Flood and J. B. Warman. (*Electronic Eng.*, vol. 28, pp. 416–421, 489–493, and 528–532; October–December, 1956.) Switching circuits using cold-cathode diodes and triodes are discussed. Circuit elements are described for performing logical operations, counting, information storage, etc. The effects of tolerances on circuit operation are examined. Applications of multicathode tubes are mentioned. 37 references.

Courtesy of PROC. IRE
and *Wireless Engineer*

681.142:538.221 57-114

High-Speed Coincident-Flux Magnetic Storage Principles—L. P. Hunter and E. W. Bauer. (*J. Appl. Phys.*, vol. 27, pp. 1257–1261; November, 1956.) Random-access magnetic-core storage systems are discussed in which the summation of coincident magnetic fluxes effects switching, the magnitude of the currents inducing the flux not being critical. Flux densities much greater than

This page has been left blank in order that readers may mount all reviews on cards.

—*The Editor*

that corresponding to the coercive field may be used, giving very short switching times.

Courtesy of Proc. IRE and *Wireless Engineer*

538.221:539.23 57-115

Resonance and Reversal Phenomena in Ferromagnetic Films—R. L. Conger and F. C. Essig. (*Phys. Rev.*, vol. 104, pp. 915–923; November 15, 1956.) Experimental results demonstrate the proportionality between magnetization reversal time and magnetic-resonance-absorption line width.

Courtesy of Proc. IRE and *Wireless Engineer*

537.226/.228.1 57-116

Ferroelectric Materials—P. Popper. (*J. Inst. Elect. Eng.*, vol. 2, pp. 450–457; August, 1956.) A short review describing the basic physical properties and indicating possible applications of ferroelectric materials, particularly $BaTiO_3$.

Courtesy of Proc. IRE and *Wireless Engineer*

57-117

A New Type of Ferroelectric Shift Register—J. R. Anderson. (IRE Trans., vol. EC-5, pp. 184–191; December, 1956.) Ferroelectric shift registers having completely independent parallel or serial inputs and outputs have been designed and constructed. The principal components of these shift registers are single crystals of barium titanate and silicon junction diodes. Two ferroelectric units and two to three silicon junction diodes are required for each stage of the shift register. Practical operating speeds for 10-stage shift registers with transistor drives are at present from 0 to 5 kc. The small size of the ferroelectric units and the low power consumption in this speed range make the ferroelectric shift register attractive for many digital circuit applications.

Courtesy of Proc. IRE

537.226/.228.1:546.431.824-31 57-118

The Effect of the Polarization Conditions on the Piezoelectric Properties of Barium Titanate—S. V. Bogdanov, B. M. Vul, and R. Ya. Razbash. (*Zh. Tekh. Fiz.*, vol. 26, pp. 958–962; May, 1956.) Polycrystalline $BaTiO_3$ elements must be polarized in a strong dc field. With specimens of appreciable thickness (15–25 mm), fields of the order of 30–50 kv/cm have to be used, with the attendant inconvenience and possible damage to the specimen. Lower polarizing voltages can be used if the temperature at which polarization is carried out is raised; the dielectric strength is not affected (Vul, *et al.*, *Zh. Eksp. Teor. Fiz.*, vol. 20, pp. 465–470; May, 1950). This has been confirmed experimentally; but even at temperatures approaching the Curie point, the polarizing voltage should not be below 5 kv/cm.

Courtesy of Proc. IRE and *Wireless Engineer*

537.227 57-119

Ferroelectricity of Glycine Sulphate—B. T. Matthias, C. E. Miller, and J. P. Remeika. (*Phys. Rev.*, vol. 104, pp. 849–850; November 1, 1956.) Glycine sulphate and its isomorphous selenate are ferroelectric with Curie points 47°C and 22°C, respectively. For glycine sulphate at room temperature the spontaneous polarization is 2.2×10^{-6} coul/cm^2 and the coercive field is 220 v/cm. A formal similarity among ferroelectric sulphates regardless of crystal structure is suggested.

Courtesy of Proc. IRE and *Wireless Engineer*

537.266/.227 57-120

Thiourea, a New Ferroelectric—A. L. Solomon. (*Phys. Rev.*, vol. 104, p. 1191; November 15, 1956.) Crystals are orthorhombic at room temperature. With electrodes on (010) faces, a pronounced dielectric anomaly is found at −104.8°C. The coercive field is less than 1000 v/cm at 60 cps and −110°C.

Courtesy of Proc. IRE and *Wireless Engineer*

535.376 57-121

Electroluminescence—D. W. G. Ballentyne. (*Wireless World*, vol. 63, pp. 128–132; March, 1957.) Sustained emission of light can be obtained from an unexcited phosphor by suspending it in the dielectric of a capacitor to which an alternating field is applied. Recent progress is described and possible applications to illumination, television, and storage of binary-code numerical data are discussed briefly. For full paper, see *Marconi Rev.*, vol. 19, pp. 160–175; 4th Quarter, 1956.

Courtesy of Proc. IRE and *Wireless Engineer*

57-122

Quarterly Report No. 1, Third Series—J. R. Bowman, C. H. T. Wilkins, *et al.* (*Quart. Rep. Computer Components Fellowship Mellon Inst.*, vii–4 pp.; October 1, 1956 to December 31, 1956.) The initial section contains some notes relative to the present experimental program and objectives of the Computer Components Fellowship, sponsored by the Air Force Cambridge Research Center under Contract No. AF 19(604)-1959. The concept of "modular printed circuitry" which appears to offer great promise for future development, has been singled out for concentrated study and evaluation. The preparation and testing of these circuits using several techniques developed by the Fellowship are described. Resistors and capacitors have been printed successfully using the screen process; results of tests at temperatures from room temperature to 260° C are presented. Analysis of a vacuum-deposited gold-palladium resistor film closely approximated that of the bulk material; a comparison of test values for resistors and capacitors deposited on alumina and on glass substrates is made. Some of the problems encountered with filaments and crucibles during vacuum evaporation are discussed; the use of boron nitride as a high-temperature crucible material appears promising. Refractory masks, formed by a casting process and by machining, were found to yield excellent resolution in the spray-painting of stannic oxide resistors; the spraying of conductive and dielectric enamels is proposed. Procedures used in the testing of components at elevated temperatures are described.

C. H. T. Wilkins

DIGITAL SYSTEM RESEARCH

57-123

Micro-Programming—R. J. Mercer. (*J. Assoc. Comp. Mach.*, vol. 4, pp. 157–171; April, 1957.) Breathes there a sophisticated programmer with soul so dead who never to himself has said: "If only I could make one little change in the order code of this machine." The basic concept of micro-programming is that every programmer should be allowed to function as his own logical designer. For, as Mr. Mercer writes: "The machine that provides the ideal order code for every problem will be the true 'general purpose' machine." Since all digital computer operations are combinative functions of a limited number of basic "micro-operations" or "sub-commands," it is the author's thesis that programmers should be allowed to work directly with these elementary operations. In support of this, he provides an analysis of the relative machine times involved in solving a cyclic permutation problem on the general purpose computer SWAC and on a hypothetical computer that allows for micro-programming. The micro-program computer solves the problem in about one-third of the time required on SWAC. A total of 46 program instructions are required for the micro-program machine as opposed to 55 of SWAC's four-address instructions. Perhaps a more complete measure of comparison might have been available if Mr. Mercer had also included a statement of the respective programming times. It is, of course, in the area of actual program construction that the concept of micro-programming breaks down as a practicable technique. This is a fact which some of our more senior programmers tend to overlook in their understandable enthusiasm for newer and more challenging methods. Since each "general purpose" order may perhaps exist as a combination of some twenty micro-operations, each of which must be performed in a rigid time sequence, the inherent difficulty of constructing and debugging a micro-program would seem to far outweigh the advantages obtained through somewhat more effective machine operation.

Herbert T. Glantz

57-124

A Method for the Evaluation of a System of Boolean Algebraic Equations—J. A. Postley. (*MTAC*, vol. 9, pp. 5–8; 1955.) The author describes a way of using punched cards to assist in making certain calculations involving a fairly large number of variables which are always either 0 or 1. The basic principle is that if one indicates the fulfilment of a condition in a given instance by punching a hole in a definite position in a card for that instance, then the simultaneous fulfillment of the condition in all the instances can be tested by sighting through the batch of cards. This idea is applied to calculating the y_r from the x_r when we have equations of the form

$$y_r = \phi(x_r, f_r(x_1, \cdots x_n), g_r(x_1, \cdots, x_n))$$

This page has been left blank in order that readers may mount all reviews on cards.
—*The Editor*

where ϕ, f_r, g_r are Boolean functions in alternative (disjunctive) normal form.

H. B. Curry
Courtesy of *Mathematical Reviews*

57-125

Digital Computational Methods in Symbolic Logic, with Examples in Biochemistry—Robert S. Ledley. (*Proc. Nat. Acad. Sci. U.S.A.*, vol. 41, pp. 498–511, errata, 796; 1955.) In two previous papers [*Nat. Bur. Standards Rep. 3363*; 1954. *J. Operations Res. Soc. Amer.*, vol. 2, pp. 249–274; 1954 (see 57-21, March, 1957)], the author proposed a technique useful for the mechanical solution of problems in Boolean algebra. The principle of this technique is that an element in a free Boolean algebra with m generators can be represented as a binary number with 2^m digits, these digits being the coefficients in a "development" (in the sense of Schröder) in which the atoms are arranged in some standard order. The present paper contains some improvements in his methods. The procedures for solving Boolean equations are presented in terms of Boolean matrices, and consideration is given to cases where the generators are subject to conditions ("constraints"). At the end the author discusses two interesting and nontrivial examples from biochemistry. It now seems clear that the author's methods save some labor in the manipulation of complex equations; but his treatment is marred by a propensity to make simple things appear difficult, and by peculiarities of notation (such as numbering his digits from right to left and his rows from the bottom upwards) which conceal the relationship of his methods to those which are more familiar.

H. B. Curry
Courtesy of *Mathematical Reviews*

57-126

Symbolic Methods in the Design of Delay- and Cycle-Free Logical Nets—George W. Patterson. (1954 IRE Convention Record, Part IV, pp. 58–64.) The author describes the relationship between the logical design of machines and the propositional calculus via truth tables. The use of the calculus in the analysis, synthesis, and optimization of logical nets is described in several examples.

S. Gorn
Courtesy of *Mathematical Reviews*

57-127

A Note on Universal Turing Machines—M. D. Davis. ("Automata Studies," in "Annals of Mathematics Studies," no. 34, pp. 167–175, Princeton University Press, Princeton, N.J.; 1956.) The author defines a universal Turing machine as one for which the set of Gödel numbers of the instantaneous descriptions is complete. By instantaneous descriptions he means initial tapes and states leading to a terminating machine reaction, and by complete he means a recursively enumerable set of integers into which every recursively enumerable set may be mapped by a suitable recursive function. This definition of a universal machine is justified by showing that 1) any computation may be encoded for such a machine, 2) any machine which may be encoded to produce any recursively enumerable set is such a machine, and 3) the encoding in 1) may be achieved by a machine whose set of Gödel numbers for the instantaneous descriptions is not merely not complete but is actually recursively enumerable.

S. Gorn
Courtesy of *Mathematical Reviews*

57-128

A Universal Turing Machine with Two Internal States—Claude E. Shannon. ("Automata Studies," in "Annals of Mathematics Studies," no. 34, pp. 157–165, Princeton University Press, Princeton, N.J.; 1956.) The author shows that any Turing machine with an alphabet of m letters and n states is "essentially equivalent" to a machine with $4mn+m$ symbols and two states. Thus a universal Turing machine may be constructed with only two states. It is also shown that a single state machine cannot compute the successive digits of an irrational number. Thus a universal machine must have more than one state. Finally it is shown that any machine with n states and an alphabet of m letters is equivalent to a machine with an alphabet of two letters and $(2^l-1)n$ states, where $l=1+[\log_2 m]$.

S. Gorn
Courtesy of *Mathematical Reviews*

57-129

The Inverstion of Functions Defined by Turing Machines—John McCarthy. ("Automata Studies," in "Annals of Mathematics Studies," no. 34, pp. 177–181, Princeton University Press, Princeton, N.J.; 1956.) This is an informal discussion of the problem of designing a machine to solve "well-defined" intellectual problems. The author identifies this problem with that of inverting a partial recursive function $f_m(n)$ to yield a function $g(m, r)$ such that $f_m(g(m, r))=r$.

S. Gorn
Courtesy of *Mathematical Reviews*

57-130

Computability by Probabilistic Machines—K. de Leeuw, E. F. Moore, C. E. Shannon, and N. Shapiro. ("Automata Studies," in "Annals of Mathematics Studies," no. 34, pp. 183–212, Princeton University Press, Princeton, N.J.; 1956.) The authors consider machines fulfilling the following conditions. 1) For every input composed of a finite binary sequence there is an output composed of a finite sequence of symbols chosen from a set S. 2) The functional relationship between the input and output sequences is computable. 3) The output corresponding to an initial segment of any given input will be an initial segment of the output corresponding to that given input. For such machines, then, there is a computable function $f(A)$ whose range is a subset of the set of infinite sequences chosen from S, and whose domain is the set of infinite binary sequences A. Any such machine together with a fixed input sequence A is called an A machine. The set of output symbols, S, of an A machine is called A enumerable. The recursively enumerable sets are those which are A enumerable for A the sequence containing 1 exclusively; the authors also call them 1 enumerable, and the enumerating machine a 1 machine. If the input tape is produced by a random device which yields the bits independently and with probability p of yielding a 1, the combination of the random device and the machine is called a p machine. S is called strongly p enumerable if there is a p machine producing S in any order with nonzero probability. If S_m is the set of output symbols printed by a p machine with a probability greater than $\frac{1}{2}$, then S_m is called p enumerable. Let A_p be the binary expansion of p, where $0<p<1$. Then the authors show that the concepts "p enumerable," "strongly p enumerable," and "A_p enumerable," are effectively equivalent. Furthermore, if p is computable, any p enumerable or strongly p enumerable set is 1 enumerable, whereas if p is not computable there are p enumerable and strongly p enumerable sets which are not 1 enumerable. Because of condition 3 above, these concepts and results extend readily to the case where S is a fixed sequence of symbols. By a "stochastic machine" the authors mean one with an initial state X_0, a countable sequence of states, $X_1, X_2, \cdots$ a countable set of output symbols $s_1, s_2 \cdots$, and a (not necessarily computable) transition probability function,

$$P(X_{i_1}, \cdots, X_{i_n}; X_p; S_q)$$

giving the probability that the machine, having gone through states X_{i_1} to X_{i_n}, will print s_q and go to state X_p. If P is computable, the machine is called a computable stochastic or c-s machine. To any p machine M one may associate an "equivalent" stochastic machine. It is a c-s machine if and only if p is computable. The authors show, finally, that every c-s machine is effectively equivalent to a $\frac{1}{2}$ machine, and every c-s-enumerable set is effectively 1 enumerable.

S. Gorn
Courtesy of *Mathematical Reviews*

57-131

Algorithmes de Markov et Théorie des Machines—Jacques Riguet. (*C. R. Acad. Sci. Paris*, vol. 242, pp. 435–437; 1956.) The concept of "normal algorithm" in the sense of A. A. Markov (*Trudy Mat. Inst. Steklov*, 42; 1954. MR 17, 1038) is related to the author's general theory (*C. R. Acad. Sci. Paris*, 237, pp. 425–427; 1953. MR 15, 559) of Ashby machines by an abstract "flow diagram."

G. Birkhoff
Courtesy of *Mathematical Reviews*

DIGITAL EQUIPMENT

57-132

Punched-Card Transcriber for Automatic Computers—Anonymous. (*Nat. Bur. Standards Tech. News Bull.*, vol. 41, pp. 38–39; March, 1957.) Punched cards are fed endwise at up to 600 cards per minute, and the columns are converted sequentially into 4- or 6-digit binary codes and recorded upon magnetic wire that can be used for input to SEAC.

R. D. Elbourn

57-133

FOSDIC II Reads Microfilmed Punched Cards—Anonymous. (*Nat. Bur. Standards Tech. News Bull.*, vol. 40, pp. 72–74; May,

This page has been left blank in order that readers may mount all reviews on cards.
—*The Editor*

1957.) Up to 13,000 punched cards are recorded upon a 100-foot roll of 16-mm microfilm. The film is searched at over 4000 cards per minute by electronic flying-spot scanning of any selected 10 card columns. Whenever those match a plugboard specification, a complete duplicate card is punched, up to a punching rate of 100 cards per minute.

R. D. Elbourn

57-134

A Tank Farm Data Reduction System—D. J. Gimpel. (IRE Trans., vol. IE-4, pp. 94–100; March, 1957.) A tank farm data reduction system has been developed for the new Tidewater Oil Company installation in Delaware by the Armour Research Foundation and Panellit, Inc. The function of the unit is to secure the temperature corrected volume of fluid in each of the approximately 100 tanks in the field. The inputs to the system are the fluid height and average tank temperature. Fluid volume is tabulated digitally as a function of the height on magnetic tape. The system automatically searches the tape for the indicated volume and multiplies the number by the temperature correction factor. This paper describes the operation of the multiplier, the tape search elements, and the sensing instruments employed in the field. The factors governing the selection of the specific elements in the storage and computing system are also discussed.

Courtesy of Proc. IRE

57-135

The IBM 705 EDPM Memory System—R. E. Merwin. (IRE Trans., vol. EC-5, pp. 219–223; December, 1956.) The IBM 705 memory system utilizes magnetic cores both as a storage element and also in a matrix address selection system. The magnetic core has been established as a memory element for large data processing machines. The core compares very favorably with other means of storage with respect to such factors as speed, reliability, size, cost, life, and simplicity of associated electronic circuitry. The memory consists of a main 20,000 character unit and a 512 character storage unit. Both are three-dimensional coincident current systems with the larger containing 35 planes of 4000 cores each and the other consisting of seven planes of 512 cores each. The basic memory cycle is 9 μsec long when operating with the input-output units or on internal transfer of data. When operating with the central processing unit a 17-μsec cycle is required. Data may be transferred within memory in five character blocks, and the five character instructions are transmitted to the control unit in one-memory cycle. Transfers between memory and the input-output and arithmetic units is serial by character. Use of the magnetic core matrix switch greatly reduces the electronic equipment required to drive the memory. Simplified circuitry requiring no adjustments eliminates any maintenance time required for making routine adjustments. Indefinite life of the core eliminates any replacement problem of the basic storage element itself.

Courtesy of Proc. IRE

57-136

Reliability of an Air Defense Computing System: Component Development—H. F. Heath, Jr. (IRE Trans., vol. EC-5, pp. 224–226; December, 1956.) This paper presents the general aspects of the component development program for the AN/FSQ-7 Air Defense Computer. The requirements of the system for high reliability and long life necessitate proper selection of the type of component, proper specification of the component to the manufacturer, and proper component application by the circuit design engineer. The component development program has made free use of ideas from the computer industry and the component industry.

Courtesy of Proc. IRE

57-137

Reliability of an Air Defense Computing System: Circuit Design—R. E. Nienburg. (IRE Trans., vol. EC-5, pp. 227–233; December, 1956.) Extreme reliability resulting in no unscheduled down-time and a low ratio of scheduled maintenance to operate time was the objective of the AN/FSQ-7 design program. The circuit design philosophy of this program is presented. In addition, an approach is given whereby the concept of marginal checking is applied to determine quantitatively 1) the relative reliability of computer circuits and 2) that a margin of safety consistent with the circuit design philosophy existed. An appendix is included setting forth actual examples in a qualitative manner.

Courtesy of Proc. IRE

57-138

Reliability of an Air Defense Computing System: Marginal Checking and Maintenance Programming—M. M. Astrahan and L. R. Walters. (IRE Trans., vol. EC-5, pp. 233–237; December, 1956.) Marginal checking by varying supply voltages for some time has been a means of preventive maintenance for electronic systems. Some important innovations have been employed in the marginal checking system of the AN/FSQ-7 air defense computer to give a more effective high-speed preventive maintenance technique. Completely automatic preventive maintenance testing is discussed incorporating program control of the marginal checking system.

Courtesy of Proc. IRE

UTILIZATION OF DIGITAL EQUIPMENT

681.142 57-139

Definitions of Terms for Program-Controlled Electronic Computers—(*Nachrichtentech. Z.*, vol. 9, pp. 434–436; September, 1956.) About 40 proposed terms are listed with their definition and English equivalent.

Courtesy of Proc. IRE
and *Wireless Engineer*

57-140

Pitfalls in Computation—I. A. Stegun and M. Abramowitz. (*J. Soc. Industrial and Appl. Math.*, vol. 4, pp. 207–219; December, 1956.) As the title implies, there are troublesome situations arising in computation (with digital machines of any kind) which should be avoided where possible. Under the subtitles, Elementary Pitfalls, Functional Relations, Indeterminate Forms, Series Calculations, Numerical Differentiation, and Numerical Quadrature, the authors indicate with examples how some of the singular situations, overflows, loss of significant figures, etc., can be obviated. It is noted that floating point operation is no panacea.

T. H. Southard

57-141

Principles of Programming—J. B. Ward. (*Electrical Eng.*, vol. 75, pp. 1078–1083; December, 1956.) This is an introduction to programming stored program computers. A hypothetical computer is described by giving the list of instructions it obeys, and the general organization and operation of the computer is stated. Two elementary computational problems are coded and flow diagrammed, as a means of describing the reduction of mathematical problems to a sequence of instructions. Elementary principles of programming are brought out in the development of these programs.

Harry T. Larson

57-142

Mathematical Machines.—(*Stroje na Zpracování Informací*, vol. 1, pp. 7–132; 1953. Czech., Russian and English summaries.) "The first part contains results of the research in numerical calculation methods suitable for solution of problems on the Czechoslovak automatic computer SAPO. Coding and a symbolism suitable to formulate instructions for the machine are explained. In this part a method is described how to form detailed instructions (the programming) for the machine in accordance with a given problem. The use of the method is illustrated by several examples. The second part deals with the use of Czechoslovak punched card machines for numerical solutions of mathematical problems." (From the English summary.) Chapter headings: 1) Introduction to Automatic Computation, 2) Coding Automatic Computation, 3) Preparing Instruction Nets, 4) Investigation of a Centered Optical System by Automatic Computation, 5) Solution of an Ordinary Differential Equation of Second Order by Automatic Computation, 6) Processing of Punched Cards, 7) Application of a Punched-Card Machine to the Solution of an Engineering Problem.

Courtesy of *Mathematical Reviews*

57-143

A Centralized Data Processing System—Jerome J. Dover. (1954 *Proc. Western Computer Conference*, pp. 172–183.) The author discusses the problems encountered in preparing large volumes of flight test data for processing by IBM Card Programmed Calculators, and the solutions arrived at in 1954 by the Air Force Flight Test Center at Edwards AFB. The reduction in calculating time provided even by CPC's gave rise to a serious bottleneck in the stage of reading data (from films, oscillographs, etc.), digitalization of this data, and application of corrections for instrument, calibration, and

This page has been left blank in order that readers may mount all reviews on cards.
—*The Editor*

other errors. A lengthy study of the problem was made by R. M. Parsons and Co., and the resulting proposal for a centralized data-processing system is described in some detail in this paper. Significant in the system is the introduction of the so-called "human filter," *i.e.*, the inspection by the test engineer of initially recorded information in uncorrected analog form with a view to eliminating large masses of information not requiring further processing. The influence of these decisions upon the validity of the final results requires no comment. However, this filtering was considered essential in order to avoid, to quote, "astronomical numbers of data points involving stacks of IBM cards literally miles high."

H. B. Prendergast

57-144

Problems in Acceptance Testing of Digital Computers—Paul Brock and Sybil Rock. (*J. Assoc. Comp. Mach.*, vol. 1, pp. 82–87; April, 1954.) To protect both the seller and the buyer a reasonable and efficient acceptance testing procedure is mandatory. The authors are especially concerned with programmed checks and the criteria for determining passage or failure of such tests. A statistical approach to this problem employing methods of sequential analysis is outlined. Perhaps the most enlightening section of this paper is the actual description of various tests in use during 1954.

W. P. Keating

57-145

The Use of Electronic Data Processing Systems in the Life Insurance Business.—M. E. Davis. (1953, *Proc. Eastern Joint Computer Conference*, pp. 11–17.) This article presented, in general terms, the status and philosophy of electronic data processing in life insurance, as of the end of 1953, starting with the work of the Committee of the Society of Actuaries appointed to study this area. To provide a basis for system comparisons the committee developed the Consolidated Functions Plan for administering life insurance policies in a manner suitable to any of several different electronic data-processing systems. This plan suggests a large reduction in the number of card files being kept and a combination of much work often done separately in different departments. The author discusses major activity categories generated by policy transactions, presents a case for maintaining, at least for the present, card files as primary records rather than tapes, discusses some limitations of current systems, the elements involved in acquiring a system, and some future implications. Some of the author's points follow: 1) Major benefits would come from adapting the Consolidated Function Plan with punched card electronic calculator methods. 2) Additional benefits would be associated with magnetic tape processing from primary card files. 3) Additional benefits associated with use of primary tape files are not large. 4) Large random access memory is not of crucial importance. 5) Approximately 90 per cent of total potential possible with highly automatic future devices could be realized with then existing equipment. 6) Magnetic tape systems can be gainfully introduced in particular areas without risking the entire recordkeeping system of the company. 7) Actual use of current models on day-to-day work is essential to the development of proper applications and of improved future processing systems. The author seems to believe that by far the major portion of the ultimate potential can be achieved with existing systems, and that radical changes should be introduced gradually, while obtaining experience with advanced equipment in particular areas which can be justified economically on their own merits without jeopardizing the entire company operation. This approach has by now become familiar in a number of areas other than insurance. It would be of interest to compare this report with a more recent survey in the same or in other areas. Unfortunately this reviewer is not familiar with a comparable and more recent survey. He suspects that most of the author's philosophy will stand up today, in the light of the experience added in the past few years.

George W. Brown

57-146

Digital Computers Can Aid Utilities—F. J. Maginniss. (*Electrical Eng.*, vol. 76, pp. 124–125; February, 1957.) This is a brief statement of the various ways digital computers are useful in the electrical power distribution utilities. Computers can be used to speed up calculations presently performed by other methods; examples are load forecasting and heat balance calculations. Computers are making it possible to solve problems by better methods; examples are solution of short-circuit studies, determination of loss formula coefficients, and solution of network transient problems. Finally, the speed and economics of computer operation make it practical to solve new problems; examples are the determination of critical shaft speeds in multiunit turbine generators and complete analysis of a steam piping system.

Harry T. Larson

57-147

Computers—The Key to Modern Manufacturing Scheduling—J. J. Gravel and T. F. Kavanagh. (IRE Trans., vol. IE-4, pp. 90–93; March, 1957.) Manufacturing operations with poor scheduling plans are headed for trouble. Load capacity analysis is a technique for measuring the feasibility and desirability of proposed scheduling plans. Simply stated mathematically, load capacity analysis consists of a series of multiplications and additions. However, the numerous computations in a typical problem usually take more time than can be allowed. The paper describes a special purpose analog computer specifically designed to solve this problem in a matter of minutes.

Courtesy of Proc. IRE

57-148

A Plan for Programming Analysis of Variance for General Purpose Computers—H. O. Hartley. (*Biometrics*, vol. 12, pp. 110–122; 1956.) The organization of computations for analyzing a k-factor experiment is described in detail for $k=3$, and it is shown how the analysis of other designs can be reduced to the factorial. As the author admits, in some cases the reduction might be more complicated than the calculations programmed, which are basically only sums and sums of squares.

A. S. Householder
Courtesy of *Mathematical Reviews*

57-149

A Manual for Coding Organic Compounds for Use with a Mechanized Searching System—T. R. Norton and A. Opler. (The Dow Chemical Co., Pittsburg, Calif., 56 pp.; 1953. Revised 1956.) **A Manual for Programming Computers for Use with a Mechanized System for Searching Organic Compounds**—A. Opler and T. R. Norton. (The Dow Chemical Co., Pittsburg, Calif., 23 pp.; 1956; and **New Speed to Structural Searches,** *Chem and Eng. News*, vol. 34, pp. 2812–2816; 1956.) One of the many applications of modern computers which particularly interests the chemist is the mechanization of searching the ponderous accumulation of literature in his field. It will be some time before the ultimate electronic library is achieved, but significant steps are being made and the work of Opler and Norton is one of them. They have devised a technique "For Coding Organic Compounds" and put it to practice with two recent computers, the Datatron and the IBM 701. Their purpose was to search rapidly the thousands of known organic chemical compounds for correlating physical, chemical, or biological properties with structure. The results of their efforts to date are summarized in an article in *Chemical and Engineering News* and the techniques are elaborated in the two manuals available from The Dow Chemical Company. The article, "New Speed to Structural Searches," provides a good introduction to and review of the problems and how they are being solved. The heart of the rapid searching technique is the coding system employed to translate the organic structure into numerical form. While a number of systems have been devised, the system of Opler and Norton represents a relatively complete reduction of organic structure notation to numerical form. This, in effect, makes the coding of searching and related routines a straightforward task for most computers. The coding system detailed in the first manual is built on a sequence of seven-digit numbers. There will be as many of these numbers in a sequence as there are chemical groups in the compound being coded. The system is designed to handle most types of organic compounds now known and is sufficiently flexible to accommodate some of the special classes of compounds currently omitted. The coding manual is clear, concise, and adequately stocked with examples. However, in the words of the authors, "The ultimate value of this convention (of coding) can only be shown by operating experience." The manual for programming the computer gives a general outline of the type of program which the authors prepared to process the coded compounds and to search the file. Due to the various modes of programming now employed in current computers the manual was written very generally. It is difficult to decide how general

This page has been left blank in order that readers may mount all reviews on cards.
—*The Editor*

or specific to make such a manual. From the point of view of those chemists who have done some coding, the compound notation system would become more acceptable if a few extra pages of detailed example were appended to the coding manual. Aside from that small point the manual very well outlines the structure of the programs. Opler and Norton are to be commended for their fine work in this and other applications of computers to chemical problems and The Dow Chemical Company should also be commended for making this information generally available.

F. H. Kruse
Courtesy of *Mathematical Tables and Other Aids to Computation*

57-150

Number Theory on the SWAC—Emma Lehmer. ("Numerical Analysis" in "Proceedings of Symposia on Applied Mathematics," McGraw-Hill Book Co., Inc., New York, vol. VI, pp. 103–108; 1956. Published for Amer. Math. Soc., Providence, R. I.) The author's account includes three classes of problems tried on the SWAC: 1) special computations (as the testing of Fermat primes, Mersenne primes, or difference sets); 2) testing of hypotheses (such as Riemann's, Polya's, Turan's, Fermat's) unfortunately with no counter-examples occurring; 3) research problems (computational explorations into Jaccbsthal's sums or Ramanujan's function) often leading to theorems of Lehmer.

Harvey Cohn
Courtesy of *Mathematical Reviews*

57-151

Experiments in Chess—J. Kister, P. Stein, S. Ulam, W. Walden, and M. Wells. (*J. Assoc. Comp. Mach.*, vol. 4, pp. 174–177; April, 1957.) This concise, interesting, and well written article contains a brief description of the authors' experiments in teaching the MANIAC computer to play chess. "The game which was played on the machine is not really chess but rather, so to speak, a miniature of it; we play on a 6×6 board, omitting the bishops, and with six pawns on each side." The program was constructed so that the machine looked two moves ahead; two by white and two by black. In analyzing its future position at the end of each such possible chain of four moves the machine considered not only material advantage; that is, actual pieces, but also positional advantage and mobility. The machine's normal "thinking" time per move was about 12 minutes, although at a crucial point in one game it pondered for a full half-hour. Three experimental "matches" are described. Initially, MANIAC was matched against itself with inconclusive results. In a later match against a "strong" player, from whom it received a considerable handicap, MANIAC was in a very favorable position until "the machine chose a weak continuation which enabled its opponent to lay a 3-move mating trap." In a final game MANIAC proved to be superior to a laboratory member who was quite a novice at the game. In summary, the authors feel that the play of the machine "could be compared to that of one who has average aptitude for the game and experience amounting to 20 or 30 full games played." The authors very correctly point out that, while such experiments are diverting and interesting, the main value of these developmets is in providing insight into the actual thought processes of the human brain. In particular, it is felt that these experiments may lead to some new understanding of the brain's technique for assimilating, organizing, and evaluating information.

Herbert T. Glantz

BOOK REVIEWS

57-152

Arithmetic Operations in Digital Computers—R. K. Richards. (D. Van Nostrand Co., Inc., New York, N. Y., iv+397 pp.; 1955.) So far as this reviewer is aware, this important book contains the most complete existing description of the application of Boolean algebra to the logical design of digital computers. It is already a standard handbook of computer design used with profit by design engineers, students of the logic of computers, and students of the logic of such operations as microcoding. The author prepared the contents as a set of notes to be used in connection with a course offered to engineers in the laboratories of the International Business Machines Corporation. As such it is strictly utilitarian. Its bibliography is incomplete. In the index one does not find the name of von Neumann, and the name of Claude Shannon occurs just once; the Shannon reference credits him with appearing to have pointed out the adaptability of Boolean algebra to the design of switching circuits. Similarly, the index lists thirteen references to IBM, but none to any of the groups which have formed Sperry-Rand. The reviewer finds no quarrel with this detachment and biased reference. However, the lack of complete references precludes use of the book as a reliable basis for estimating the complete state of the science at the time the book was written. The author had a goal which he met well. He wanted to describe design of computing instruments and he turned to the material at hand. Where it was convenient and where it served his purpose he took material from other sources, and to the extent that he had information readily available concerning the ultimate source he seemed to give it. Thus he gives no fewer than a dozen examples of binary adders in block diagram form (using Boolean elements) and discusses their relative merits. He goes into considerable detail concerning the difficulties of designing decimal components and again he illustrates his arguments with many block diagrams. In similar fashion he examines the other every-day problems of the designer of circuits for computers. He claims no new results and hence he feels little need for discussing relations between his results and those of other workers. There is no discussion of the practical limitations of circuit design. Thus the deterioration of pulses on transmission lines and the effects of such deterioration on the logical functioning of the circuits (with the implications which can be drawn concerning the speed of computation which is safe) are not discussed in detail. Indeed, only the most idealized engineering realizations of the Boolean elements are discussed at all. These realizations are limited to vacuum tube circuits with no discussion of logical elements built of transistors and magnetic cores. Similarly, the reduction of Boolean expressions to most economical form is discussed by implication only and no thorough analysis is attempted. These omissions are not shortcomings, for the book is a thorough exposition by example of much of the formalism which is necessary for an effective application of knowledge in these omitted fields to the design of computers. Thus, there is no better place for an engineer or other person not already well informed concerning symbolic logic and its application to switching circuits to learn how to combine the components which may be developed from new electronic devices (such as transistors) into effective computing instruments. Similarly, there is no other source which seems to give as easy access to the ideas involved in studies of fast and complex circuitry such as, for example, A. Weinberger and J. L. Smith, "A One-Microsecond Adder Using One-Megacycle Circuitry," IRE TRANS. vol. EC-5, pp. 65–73; 1956. (See 56-178, December, 1956.) This article also appears under the title, "The Logical Design of a One-Microsecond Parallel Adder Using One-Megacycle Circuitry," in 1956 *Proc. Western Joint Computer Conference*, San Francisco, Calif., sponsored by the AIEE, the Assoc. for Computing Machinery, and the IRE. Published by the AIEE, New York, pp. 103–108, 1956.) The author has included a chapter on computer organization and control and one on programming. These are fairly superficial accounts suitable for project engineers but certainly not sufficiently complete to serve as serious expositions of these complicated subjects. Presumably these chapters were added to describe the general machine to the engineer working on its components. However, the subjects treated are not completely within the scope of Boolean algebra (as the author notes on p. 339) and hence not completely within the scope of this book. In short, the author has prepared a well-directed set of notes for use as a practical handbook for anyone interested in the inside of an electronic computing instrument. The book contains no problems for solution by the reader, but otherwise it is entirely suitable for use as a textbook for a course covered by the material accorded attention. The table of contents follows: Chapter 1) "Symbolic Representation of Quantities," 2) "Boolean Algebra, Applied to Computer Components," 3) "Switching Networks," 4) "Binary Addition and Subtraction," 5) "Binary Multiplication and Division," 6) "Decimal Codes," 7) "Counting, Binary and Decimal," 8) "Decimal Addition and Subtraction," 9) "Decimal Multiplication and Division," 10) "Miscellaneous Operations," 11) "Computer Organization and Control," 12) "Programming, Bibliography, and Index."

C. B. Tompkins
Courtesy of *Mathematical Tables and other Aids to Computation*

This page has been left blank in order that readers may mount all reviews on cards.
—*The Editor*

57-153

Analog Computer Techniques—Clarence L. Johnson. (McGraw-Hill Book Co., Inc., New York, N. Y., 280 pp.; 1956.) This book was written as a textbook for a course in the operation of an electronic differential analyzer. Thus, its objectives are highly specialized; it does not treat network analyzers, electromechanical differential analyzers, or any of the other analog computers. Furthermore, the author is interested in techniques of using analog computers, and only passingly in techniques of building such computers; in particular, he limits his discussion of engineering to the information which he feels is essential in proper application of the computer. In this application, however, he is insistent that the operator "feel" the way in which the machine actually functions analogously to the operation being simulated. Thus, he would like to have the operator understand why a differential equation describes a system being studied and why the same differential equation describes the operation of the analog computer used to find a solution. By bypassing the differential equation, then, the operator understands how the analog computer simulates the system to be studied. In order to get on with his task, the author limits himself bibliographically to those works which are of almost immediate importance. It is somewhat shocking to find a book on analog computation in whose index the name of V. Bush appears only once. The Massachusetts Institute of Technology is likewise listed only once, and the name of S. H. Caldwell is not in the index. Thus, explicit reference to the great pre-war development of differential analyzer techniques at M. I.T. is almost completely lacking. However, the reviewer can see no real necessity for recalling old times in a strictly utilitarian textbook; the bibliographical omissions were mentioned above to illustrate the utilitarian character of the book and not to criticize the author's scholarship. The one departure from the restrictions to electronic differential analyzers occurs in the last chapter when incremental digital computers are discussed (the so-called digital integrating differential analyzers). The thought here is that their use is more similar to that of the electronic differential analyzers than to that of the other digital machines. The material in the text is presented in a lucid and elementary way. It is accompanied by problems for the student and numerous illustrations. Each chapter has a selected list of references, short but seemingly adequate for the material involved. The reviewer feels that the book well attains its fundamental purpose of presenting the material that must be known to a person who intends to operate a standard electronic differential analyzer effectively. The table of contents is: Chapter 1) "Introduction," 2) "The Linear Computer Components," 3) "Time- and Amplitude-Scale Factors," 4) "The Synthesis of Servomechanism Systems," 5) "Multiplying and Resolving Servos," 6) "Additional Computer Techniques," 7) "The Representation of Nonlinear Phenomena," 8) "Multipliers and Function Generators," 9) "Miscellaneous Applications of the Electric Analog Computer," 10) "Analog Computer Components and Computer Control," 11) "The Checking of Computer Results," 12) "Repetitive Analog Computers," 13) "The Digital Integrating Differential Analyzer."

C. B. Tompkins
Courtesy of *Mathematical Tables and other Aids to Computation*

57-154

Sčetno-Rešayuščie Ustroǐstva (Computing Devices)—S. O. Dobrogurskiĭ and V. K. Titov. (Gosudarstv. Izdat. Oboronnoi Promyšlennosti, Moscow, 224 pp.; 1953.) Some mechanical, electromechanical and electronic analog-computer elements are described. Included are most of the well-known standard forms: integrators, multipliers, resolvers, adders, etc. Assembly drawings of some of the elements are shown. On occasion the author dwells on such structural details as mechanical fasteners (bolts, taper pins, keys, couplings), stop mechanisms, anti-backlash, and wear-compensating devices. Sources of error in various computer elements are pointed out. In a few cases detailed error analyses are made. Typical are the discussions of statistical variation in hole and shaft diameters in manufacturing a third-class fit, derivation of formulas for mean-square deviation, and analysis of potentiometer errors.

W. W. Soroka
Courtesy of *Mathematical Reviews*

57-155

Random Processes in Automatic Control—J. Halcombe Laning, Jr. and Richard H. Battin. (McGraw-Hill Book Co., Inc., New York, N. Y., ix+434 pp.; 1956.) This book is a presentation of the techniques useful in analyzing and synthesizing control systems whose inputs are random functions of time. It contains sufficient background in statistical methods to allow the control systems engineer, starting with only a speaking acquaintance with statistics, to acquire a working knowledge of these methods. The general aim of the book is to give the engineer the tools to enable him to fashion some sort of figure of merit for a control system subject to random inputs. Two preliminary chapters devoted to an exposition of "Basic Concepts of Probability Theory" and "Descriptions of Random Processes" prepare the student (or engineer) for the analyses of actual systems. The properties of correlation functions, stationary and ergodic processes, and the energy spectral density are all treated. These chapters are interlarded with enough specific examples, illustrative of the ideas presented in the text, that the reader can gain some facility in giving analytic expression to physical statistical problems. These preliminary chapters are sufficiently comprehensive that the reader need not, in the subsequent discussions, be limited by only an intuitive grasp of probability and random processes. On the other hand, a complete assimilation of this background material is not required in order to make use of the techniques presented in the principal chapters concerned with systems engineering. In these subsequent chapters the extensive use of the mean-squared error as a performance index and its limitations are discussed candidly. In fact, the careful delineation of the state of the techniques discussed is one of the principal features of the book. For those areas for which inadequate analytical theory exists, the situation is outlined and the analog computation methods, which take the place of the undeveloped mathematical theory, are explained. Since this is the situation for systems with variable coefficients, or with nonstationary inputs, these analog computing methods are finding much use in analyzing actual control system problems such as occur in fire control system design. The time-saving method of adjoint simulation for obtaining mean-squared errors is spelled out in some detail. For systems with stationary inputs and constant coefficient control systems, mathematical means are developed for obtaining the mean-squared error in terms of the system weighting function, or transfer function, and the input spectral density. The final two chapters in the book deal with system synthesis rather than analysis. The Wiener theory for synthesizing the optimum system weighting function is developed. In the last chapter it is extended to cases for which the input signal is not stationary, and for which the past history is known over only a limited time. A well annotated bibliography is provided, and a short but complete explanation of the analog computer is given in an appendix. For either the student or the practising control systems engineer, this book offers an understandable statement of the current techniques (and their limitations) available for the analysis and synthesis of control systems subject to random input signals.

W. F. Cartwright
Courtesy of
The American Mathematical Monthly

57-156

Introduction to Numerical Analysis—F. B. Hildebrand. (McGraw-Hill Book Co., Inc., New York, N. Y., x+511 pp.; 1956.) In the preface, the author states that the book is intended to provide an introductory treatment of the fundamental processes of numerical analysis which is compatible with the expansion of the field brought about by the development of high-speed computing machines. However, he indicates that he intends to take into account the fact that very large amounts of computation will continue to be effected by desk calculators (and by hand or slide rule) and that familiarity with computation on a desk calculator is a desirable preliminary to computation on a high-speed computer. He expects that it would be possible to provide a survey of a substantial portion of the text in a single semester, and that a more thorough coverage could be provided in two semesters. It is evident that the book will be much more valuable to the user of desk computers than to the user of high-speed computers. A better background for the field of modern computation would be provided by a book which covered only the most frequently used methods of hand computation and which at the same time was written from the standpoint of high-speed computing. Until such a book appears, however, the present volume will probably find extensive use as a text for courses in numerical analysis. It is very clearly written in sufficient detail so that a good student should be able

This page has been left blank in order that readers may mount all reviews on cards.
—*The Editor*

to learn the material with little or no assistance. At the end of each chapter a large number of illustrative numerical problems are given. These should be of immense value to the instructor, since one of the drawbacks of other available texts in numerical analysis has been the lack of enough good problems. The introductory chapter contains a discussion of various types of errors, including statistical errors, and analyzes the growth of such errors. Some mathematical preliminaries are also provided. The next four chapters, covering 140 pages, contain a thorough discussion of interpolation, numerical quadrature, and numerical differentiation. Use is made of divided differences, methods based on Lagrange's interpolation formula, finite differences, and difference operators. The next chapter contains a description of the standard methods for solving ordinary differential equations together with an analysis of so-called "parasitic" solutions and of propagated errors. One section is devoted to two-point boundary value problems and another section describes methods for solving characteristic value problems. Chapter VII deals with least squares polynomial approximation with special reference to Legendre, Laguerre, Hermite, and Tchebycheff approximations. Factorial power functions and summation formulas are also discussed. The chapter is concluded by a description of several techniques for smoothing empirical data. Chapter VIII treats Gaussian quadrature and other related quadrature methods including Hermite, Legendre-Gauss, Laguerre-Gauss, Hermite-Gauss, Tchebycheff-Gauss, Jacobi-Gauss, Radan, and Tchebycheff quadrature. Chapter IX considers approximations of various types including Fourier approximation for both continuous and discrete ranges, exponential approximation, Tchebycheff interpolation, and approximation by continued fractions. The concluding chapter, which is practically independent of the preceding chapters, summarizes a number of methods for the numerical solution of sets of linear algebraic equations, nonlinear algebraic or transcendental equations, and nonlinear algebraic equations in particular. There is a bibliography with 276 references and an appendix with a directory of methods. In general, it seems that too many methods are covered and that there is not enough discrimination nor enough discussion of the relative merits of the various methods. In particular, far too much space is devoted to interpolation and to finite differences. It would certainly seem that a knowledge of only a fraction of the methods presented would enable one to handle in a satisfactory way all except the most unusual interpolation problems. Except for the section on error formulas, the entire chapter on operator methods might well have been eliminated. Similarly a large number of methods are presented for solving initial value problems for ordinary differential equations and yet only a single section is devoted to the important topic of two-point boundary value problems. The treatment of methods for solving systems of linear algebraic equations could perhaps have been more detailed, if the length of some of the earlier chapters on interpolation had been reduced. The material in Chapters VII, VIII, and IX on least squares polynomial approximation, Gaussian quadrature, and approximations, respectively, is very well presented, and does not appear to be readily accessible in as palatable a form elsewhere. By concentrating only on those methods which are most useful for actual computation, an instructor should be able to present a very satisfactory one-semester course in numerical analysis using this book, which is probably the best textbook available today for an elementary course in numerical analysis. The book should be part of the personal library of anyone working in the computing field.

David Young
Courtesy of
The American Mathematical Monthly

57-157

Numerical Analysis—Z. Kopal. (John Wiley and Sons Inc., New York, N. Y., xiv+556 pp.; 1955.) This book is intended to be a research handbook as well as a text in numerical analysis as applied to functions of a single real variable. Although elementary calculus and "some algebra" are stated to be the only really necessary prerequisites; nevertheless, it would appear that considerably more mathematical background would be needed in order to understand some of the more advanced chapters. The book will probably find wider use as a research handbook than as a text for a course in numerical analysis. It would only be possible to cover a fraction of the material in a single semester, and in a two-semester course in numerical analysis, one would almost certainly wish to include a considerable number of topics which are not treated in the book such as algebraic and transcendental equations involving a single variable, systems of equations, matrix problems, and partial differential equations. The book is written from a point of view considerably removed from that of the user of high-speed computing machines and lying somewhere between applied mathematics and hand computation. Some material is included which appears to fall in the area of classical analysis rather than numerical analysis. For instance, four pages are devoted to the standard proof of the Picard iteration process for solving ordinary differential equations and ten pages are used to treat the method of Frobenius. Since these topics are not treated specifically from the point of view of numerical computation, it would seem that only the results could be given, together with a reference to a book on ordinary differential equations. The introductory chapter provides an interesting history of number systems and of numerical analysis as well as motivations for the study of the subject. Chapter II contains descriptions of the usual methods of polynomial interpolation, including the "throwback" method of Cromie which uses modified differences and describes an interesting application to curve fitting. The next chapter contains descriptions of the standard methods for solving ordinary differential equations. In addition, some new material on "successive extrapolation" is developed and applied to equations of the form $y''+f(x)y=g(x)$. The use of adjoint systems to estimate the propagated errors of various numerical procedures is also discussed. Chapter V which is entitled "Boundary Value Problems" is primarily concerned with characteristic value problems associated with linear ordinary differential equations. By the use of finite difference methods the problem is reduced to that of finding the characteristic values of a matrix. The use of higher order difference equations and also various "extrapolations to zero grid size" are also considered. Chapter VI treats these problems by variational methods including the Rayleigh-Ritz method, Schwarz's method, a "collocation" method which is related to Lagrange's method of interpolation, and a least squares method. Chapter VII deals with a number of methods for performing numerical quadrature. Methods which use unequally spaced intervals as well as those using equally spaced intervals are discussed. The concluding chapter deals with integral equations and integro-differential equations. The following subjects are treated in the five appendices: operational approach to finite difference formulas; trigonometric intepolation and Tchebycheff polynomials; coefficients for "mechanical" quadrature formulas; and algebraic equations and systems of linear equations. In general, there seems to be too much attention devoted to some of the topics with the result that the scope is much narrower than one would expect from a book of this length. The book is very clearly written in a pleasing style and the arguments are easy to follow. The chapters on variational methods for solving characteristic value problems and on integral equations are particularly interesting. At the end of each chapter a fair number of examples are provided together with supplementary notes and some research problems. The book is not entirely devoid of misprints; for instance on page 20, line 27, $a_1, a_2, \cdots, a_n$, should be $a_0, a_1, a_2, \cdots, a_n$ and on page 21 formula (II-B-4) should have $p_n^I(a_j)$ in the denominator. Although this book is not recommended as a text to be used in a course in numerical analysis, it contains a great deal of interesting material and should prove very useful to those who work in the field of numerical analysis.

David Young
Courtesy of
The American Mathematical Monthly

57-158

An Introduction to Junction Transistor Theory—R. D. Middlebrook. (John Wiley and Sons, Inc., New York, N. Y., xxiv+295 pp.; 1957.) The author succeeded in filling a gap in the transistor literature by creating a book for electronic engineers who are interested in the physical aspects of transistor operation. It forms a bridge between the physics of semiconductors and the circuit properties of junction transistors. This book provides a continuous development of basic junction transistor theory, starting with a thorough qualitative analysis of the most important fundamental physical principles and ending with the derivation of a new high-frequency equivalent circuit. The standard knowledge of mathematics, physics, and electronics of an electronics engineer is sufficient for understanding this book. The main chapters of the book deal with: qualitative development of junction transis-

This page has been left blank in order that readers may mount all reviews on cards.
—*The Editor*

tor theory; qualitative and quantitative semiconductor physics; current flow in semiconductors; boundary values for forward biased *p-n* junctions (energy diagram, carrier distribution, etc.); *p-n* junctions under applied dc potential; *p-n-p* transistors (internal capacitance and feedback effects); small-signal ac equivalent circuits (complete and simplified configuration for circuit applications). The qualitative explanations are very clear and easy to read. The quantitative analysis is straightforward and the mathematical formulas are carefully derived and discussed. Repetitions on purpose are helpful in learning the rather complicated principles of transistor action. Many figures clarify the physical meaning of mathematical derivations. A new high-frequency equivalent circuit, derived from the physics of the transistor and proven experimentally up to about twice the cutoff frequency, is a major contribution. Only the classical references are mentioned in each chapter. It is felt that more experimental results should have been given to provide a comparison with the calculations. The very important topics of high injection and surface effects are treated rather briefly. Also, the drift transistor is not mentioned, but the modest title of the book "An Introduction to . . . " tells the reader that this book does not intend to be a sort of a "handbook on transistor theory." A discussion of many of the formulas with respect to silicon devices (not just germanium ones) would have been desirable. This book is well written and the transistor theory discussed in a surprisingly clear, logical, and concrete manner. The electronics engineer and the student who is not satisfied with transistor knowledge from the "black box" point of view will gain a thorough understanding of basic transistor problems (the "hows" and "whys") by studying the pages of this book.

Peter Kaufmann
Courtesy of PROC. IRE

57-159

Electronic Analog Computers, Second Edition—Granino A. Korn and Theresa M. Korn. (McGraw-Hill Book Co., New York, N. Y., xiv+452 pp.; 1956.) The second edition of this book covers the same range of material as the first edition. (See 53-47, September, 1953.) The various chapters have been expanded and brought up to date. In particular, there are more examples, and more attention is paid to nonlinear situations and their representation by diode circuits. The reference quality of the book has been improved by the addition of Tables of Transfer Impedances for Operational Amplifiers. A table giving circuits for the generation of functions of time as solutions of differential equations, and a table of potentiometer circuits of function generation are also given. For those interested in analog computation and analog computers this book is a necessity.

H. D. Huskey